UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

The first step is the hardest: A mixed methods single-case experimental design study of a vr-enhanced training program in a forensic youth care setting

Klein Schaarsberg, Renée. E.; Klinkhamer, Nicole; van Dam, L.; Bouwmeester, Samantha; Lindauer, Ramón J. L.; Popma, Arne

DOI
10.1007/s10802-025-01313-1

Publication date
2025

Document Version
Final published version

Published in
Research on Child and Adolescent Psychopathology

License
CC BY

Link to publication

Citation for published version (APA):
Klein Schaarsberg, R. E., Klinkhamer, N., van Dam, L., Bouwmeester, S., Lindauer, R. J. L., & Popma, A. (2025). The first step is the hardest: A mixed methods single-case experimental design study of a vr-enhanced training program in a forensic youth care setting. *Research on Child and Adolescent Psychopathology*, *53*(12), 1733-2753. https://doi.org/10.1007/s10802-025-01313-1

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.
UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:15 Nov 2026

Research on Child and Adolescent Psychopathology (2025) 53:1733–1753
https://doi.org/10.1007/s10802-025-01313-1

The First Step is the Hardest: A Mixed Methods Single-Case Experimental Design Study of a VR-Enhanced Training Program in a Forensic Youth Care Setting

Renée E. Klein Schaarsberg[1,6] · Nicole Klinkhamer[2] · Levi van Dam[2,3] · Samantha Bouwmeester[7,8] · Ramón J. L. Lindauer[4,5,6] · Arne Popma[1,5,6]

Accepted: 11 March 2025 / Published online: 14 April 2025
© The Author(s) 2025

Abstract

Street Temptations is a virtual reality-enhanced training program developed as an add-on intervention for adolescents with disruptive behavior problems. Using mentalization as a primary treatment mechanism, the program aims to address cognitive distortions and increase motivation for behavior change. This study provides an initial evaluation of Street Temptations in a forensic youth care setting, utilizing both quantitative and qualitative data within an A-B-A' single-case experimental design. Throughout a baseline, intervention, and follow-up phase, adolescents completed daily measurements on cognitive distortions and motivation for behavior change. Secondary outcomes were assessed before baseline, after intervention, and after follow-up, which included mentalization and perspective-taking. Qualitative data were collected after the intervention through separate interviews with adolescents and Street Temptations therapists. A total of 8 adolescents were included in the study, 5 of whom dropped out, leaving 3 who provided sufficient data for visual inspection of the outcomes. The observed patterns suggested an increase in adolescents' awareness of others' perspectives during participation in Street Temptations, potentially indicating a positive effect on cognitive distortions. However, daily assessments did not show an increase in adolescents' motivation for behavior change. Positive effects of Street Temptations and virtual reality, as well as areas for improvement, were highlighted in all interviews. Recommendations for further development and implementation are discussed. Overall, this study provides preliminary support for the use of Street Temptations as an add-on intervention in forensic youth care.

Trial registration number

Central Committee on Research Involving Human Subjects NL75545.029.20. Netherlands Trial Register NL9639; https://www.trialregister.nl/trial/9639 (30.07.2021).

Keywords Mentalization · Virtual reality · Single-case experimental design · Adolescence · Forensic youth care

✉ Renée E. Klein Schaarsberg
r.e.kleinschaarsberg@amsterdamumc.nl

[1] Child and Adolescent Psychiatry & Psychosocial Care, Amsterdam UMC, Vrije Universiteit Amsterdam, Amsterdam, The Netherlands

[2] Department of Child Development and Education, University of Amsterdam, Amsterdam, The Netherlands

[3] Dutch Innovation Network for Societal Youth Challenges, Garage2020, Amsterdam, The Netherlands

[4] Child and Adolescent Psychiatry, Amsterdam UMC, University of Amsterdam, Amsterdam, The Netherlands

[5] Levvel, Academic Center for Child and Adolescent Psychiatry, Amsterdam, The Netherlands

[6] Amsterdam Public Health, Mental Health, Amsterdam, The Netherlands

[7] Out of the Boxplot, Rotterdam, The Netherlands

[8] Department of Developmental Psychology, Tilburg School of Behavioral Sciences, Tilburg University, Tilburg, The Netherlands

Introduction

In the beginning of the 20th century, the *Oxford English Dictionary* introduced the word 'mentalize' (Allen et al., 2008). Since then, the construct has grown to be known as a crucial ability to navigate human social life (Gergeley et al., 2002). Mentalization, as a mental process, enables us to interpret behavior of both ourself and others as meaningful, using intentional mental states, e.g. thoughts, feelings, and motives, as a basis for such interpretations (Allen et al., 2008; Bateman & Fonagy, 2004).

Since its first introduction, a substantial amount of work has been focused on discovering how mentalization is related to a variety of mental health problems and how mentalization-based treatment approaches may help to alleviate such problems. Mentalization is operationalized as reflective functioning, which may help to understand the nature of mental health problems that can be related to mentalization. For example, when encountering issues in mentalization, i.e. reflective functioning, adequately interpreting others' motives and behavior becomes difficult (Baglio et al., 2016; Happé, 1994; Yirmiya et al., 1998). Furthermore, problems regarding self-awareness and -regulation can be expected (Gergeley et al., 2002). As a result, misunderstanding social cues becomes more likely and interpersonal communication can become more impulsive (Möller et al., 2014). This misinterpretation of social information, through disturbance of a social cognitive process like mentalization, may result in aggressive or antisocial behavior (Arsenio, 2010; De Castro & Van Dijk, 2017).

Mentalization may function as both a possible protective factor, as well as a potential risk factor for such disruptive behavior. Consequently, to mitigate this behavior, focusing on the enhancement of mentalization may be a relevant target in therapy (Klein Schaarsberg et al., under review). At the same time, mentalization has been found to function as a prerequisite to engage in and benefit from psychotherapy, also significantly predicting premature treatment dropout (Allen et al., 2008; Jørgensen et al., 2021).

Although still in its infancy, the relevance of mentalization and mentalization-based treatment approaches particularly in relation to youth with (serious) disruptive behavior problems are gaining interest. By means of the current study we hope to instigate this interest, as mentalization presents itself as a promising but also understudied concept in relation to this population (Kasper et al., 2024). More specifically, to explore the added value of a mentalization-based approach to forensic child and youth care, Street Temptations was developed and studied.

Street Temptations

Street Temptations (ST) is an add-on intervention for adolescents with disruptive behavior problems. Adolescents engage in perspective-taking exercises using fictional and real-life scenarios. They create backstories for the characters involved and reflect on connections between observable behavior and internal mental states. This reflective process is the essence of mentalization, as previously explained (Choi-Kain & Gunderson, 2008; Fonagy et al., 1998).

An innovative strength is the use of virtual reality (VR). First, VR aids perspective-taking exercises by incorporating visual elements into therapy, reducing cognitive demands for creating mental representations of situations that are often used for such exercises (Cornet & van Gelder, 2020; van Gelder et al., 2019). This makes VR especially beneficial for ST. Second, this visual support is likely to benefit (part of) the target population, including adolescents with mild-to-borderline intellectual disabilities, who may require visual aids for effective treatment (de Wit et al., 2023; Moonen & Kaal, 2017; Vermaes et al., 2014). Third, VR may also appeal to a tech-savvy generation (Lochman et al., 2019; Weisz et al., 2018), while providing a safe yet realistic environment to challenge perspective-taking in social situations involving disruptive behavior (Bailenson, 2018). Fourth, adolescents may have difficulty engaging in therapeutic conversations (O'Keeffe et al., 2018; Stige et al., 2021). Using VR could help expedite the conversation between therapist and adolescent, having not only to rely on talking strategies. ST's VR content, and the rest of the program, have been developed in close collaboration with adolescents and practitioners (Klein Schaarsberg et al., 2023, 2024).

By using mentalization as an important mechanism, ST's primary aim is to challenge adolescents to reflect on both their own and others' behavior and underlying mental states. In doing so, focus lies on the reinforcement of adolescents' treatment responsivity, as not all adolescents with disruptive behavior problems seem to profit from existing evidence-based treatment options (Riise et al., 2021). More specifically, it is aimed to enhance two particular treatment responsivity indicators, being motivation for behavior change and self-serving cognitive distortions.

Treatment Responsivity Indicators

Motivation for behavior change is an important factor for treatment responsiveness (Miller & Rollnick, 2002; Olver et al., 2011). However, in the case of adolescent disruptive behavior, intrinsic motivation for behavior change can be particularly low (Brauers et al., 2016; Englebrecht et al., 2008; Harder et al., 2012; Hillege et al., 2017). Therefore,

addressing motivation appears vital for sustainable behavioral changes and treatment success (Carl et al., 2020; Harder, 2018).

To foster motivation for behavior change, one must recognize problematic outcomes of certain behavior, desire different behavior, and be open to seeking help in this instance (Drieschner et al., 2004; Ward et al., 2004). To achieve this, a certain level of self-reflection and insight is required (te Velde et al., 2015). For that reason, as self-reflection and insight are part of mentalization, ST is assumed to be of added value for the enhancement of adolescents' motivation. Through exercises that challenge adolescents to explicitly consider mental states while attributing meaning to certain behavior, it is attempted to create space for the evaluation of one's own interpretations and behavior. In that way, ST intends to challenge adolescents to identify problematic outcomes of their own interpretations and behavior, for example, fostering the motivation to change this behavior.

Similarly, such exercises are assumed to be of added value in addressing cognitive distortions. Such self-serving distortions are described as inaccuracies or biases in relation to thoughts, for example, which are used to attribute meaning to behavior. An example is a tendency to interpret ambiguous behavior of someone as hostile, potentially leading to a hostile, inadequate reaction in return (Barriga & Gibbs, 1996; Barriga et al., 2000, 2001, 2008). A negative association between cognitive distortions and treatment response for treatment of aggressive behavior has been found (Smeijers et al., 2018), stating the importance of addressing cognitive distortions in addition to motivation, for instance. It is expected that ST, through taking on others' perspectives and actively attributing meaning to behavior, will help to challenge possible inaccuracies and biases in this process of attributing meaning as well (Gibbs et al., 2013). In other words, during the ST exercises therapists can help adolescents to see a different perspective, for example, and to identify other interpretations of certain behavior.

Potential Treatment Outcome

Advancing on the expectation explained in the previous paragraph, it may be relevant to take emotion regulation strategies into account as a preliminary treatment outcome of participating in ST. Negative distortions, like the ones discussed, have been found to positively correlate with usage of non-adaptive regulation strategies, for example (Deperrois & Combalbert, 2022). Furthermore, emotion regulation is assumed to be highly relevant in relation to disruptive behavior (Njardvik et al., 2022; te Brinke et al., 2021). Although ST's focus is not on teaching different emotion regulation skills per se, its primary focus on advancing reflection and perspective taking could translate into improvement of one specific emotion regulation strategy: "reappraisal." This strategy applies to the meaning or self-relevance that is attributed to a certain situation (Gross, 1998, 2015). Therefore, it was deemed relevant to discover how ST's focus on reflection and perspective taking, and the application thereof through the regulation strategy "reappraisal", related to a participant's perceived control over (impulsive) behavior in moments of high arousal.

The Current Study

The purpose of the present study was to provide a first and thorough exploration of ST's added value as an innovative VR-based add-on intervention for youth with (serious) disruptive behavior problems. The main focus was on the effects of a mentalization-based approach on motivation for behavior change and cognitive distortions, as important treatment responsivity indicators, as well as on additional effects on emotion regulation as a potential treatment outcome. These effects were studied by applying a single-case experimental design (SCED) in two research settings, combined with additional quantitative measures as well as a qualitative evaluation of the intervention and the VR component more specifically.

A SCED study, involving repeated measurements on individual subjects before, during, and after participating in an intervention, allows individuals to serve as their own controls (Krasny-Pacini & Evans, 2018). Consequently, a SCED study ideally fits clinical research involving small and heterogeneous populations that are difficult to study using standard group designs like the randomized controlled trial. The population targeted by ST fits this description. Additionally, extensively studying a few participants enhances the understanding of individual cases and enables the discovery of individual intervention effects amidst performance variability (Normand, 2016). Altogether, these characteristics ensure that a SCED seems especially suited for this current study. By conducting our SCED study, we aimed to provide valuable insights into ST as an add-on intervention and contribute to the advancement of forensic youth care.

Methods

The protocol for this study has been published and will be referenced for additional information where appropriate (Klein Schaarsberg et al., 2022). We followed the Single-Case Reporting Guideline in BEhavioural Interventions (SCRIBE) 2016 for the reporting of our single case study (Tate, Perdices, Rosenkoetter, Shadish, Tate et al., 2016; Tate, Perdices, Rosenkoetter, Shadish et al., 2016).

Design

A non-randomized, non-concurrent, multiple-baseline design with replication across single participants was applied (see Fig. 1). Randomization across different baseline lengths was not performed, as baseline lengths would inherently vary due to natural circumstances in clinical practice. Moreover, mandatory waiting would most likely increase the already substantial risk of dropout within this population. Additionally, baselines would naturally be non-concurrent because it was impossible to assign all participants to a therapist at the same time, resulting in varying time lapses between phases across participants in any case. Blinding of participants, ST therapists, and researchers to any component of the study was not used, because this was not feasible due to the nature of the study.

Participants were asked to complete daily repeated measurements during a baseline phase (phase A), an intervention phase (phase B), and a post intervention follow-up phase (phase A'). The baseline phase functioned as a control and was therefore compared with the intervention and follow-up phases. The baseline phase lasted a minimum of 7 days, to keep the baseline length as short as possible to limit the burden on participants while still getting a representative estimate of the outcomes at baseline. The exact start date of the intervention depended on the participants' presence and the ST therapists' schedules. The intervention phase lasted at least four weeks, accommodating two ST sessions of 45–60 min per week. When this pace was not achieved, for example, because participants did not attend appointments, the intervention phase was extended. After completing all ST sessions, participants entered the follow-up phase, lasting a final 7 days.

Secondary outcomes were assessed by means of pre-, post-, and follow-up measurements. In addition, separate interviews with participants and ST therapists were performed at the end of the intervention phase. Lastly, 3 and 6 months after ending the study period, participants were approached again for an interview, to evaluate potential long-term effects.

Procedural Changes

Certain amendments to the original protocol have been made before and during the current study. These amendments followed both from practicalities as from our evolving understanding of the research design in combination with clinical practice. For a detailed description of the amendments, please see Supplementary Material 1.

Selection Criteria

Adolescents had to meet the following criteria to be eligible for inclusion, assessed by means of clinical judgment: (1) aged between 12 and 18 years, (2) antisocial or externalizing behavioral problems, (3) deficits regarding cognitive distortions or treatment motivation, (4) presence or risk of delinquent behavior, (5) assigned to ST after multidisciplinary consultation, (6) expected stay of at least 2 months, and (7) basic understanding of mobile apps. Clinical judgment, rather than standardized assessment, was chosen to reduce the burden on adolescents and clinicians. Additionally, creating a homogeneous sample—often achievable through standardized assessments—was not a goal of this study. Instead, we deliberately employed a SCED to accommodate the inherent heterogeneity of the population studied. A potential participant who met any of the following exclusion criteria was excluded from participation: (1) severe physical impairment such as deafness or blindness, (2) severe psychiatric problems such as psychosis or high

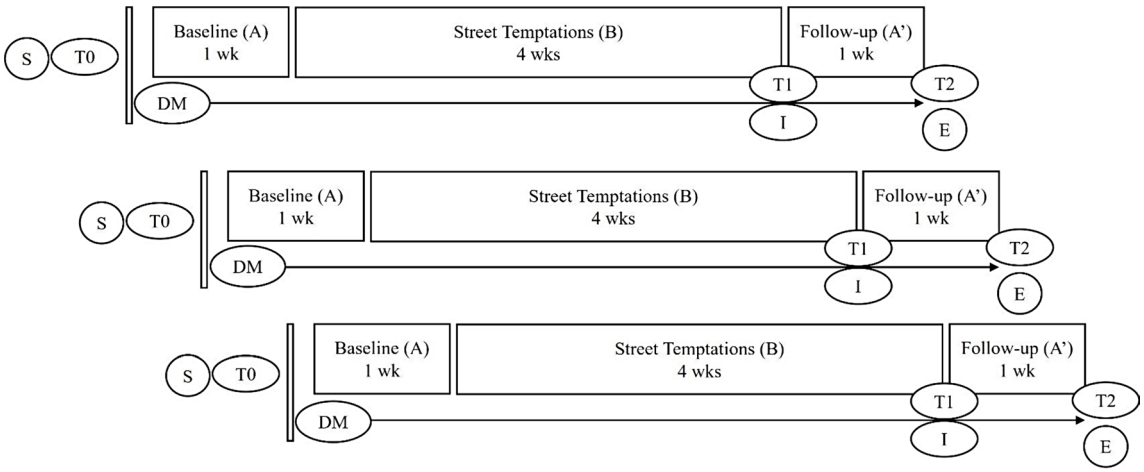

Fig. 1 Overview of the study design for three cases. Note. The daily repeated measurements start directly after T0, on the same day; DM = daily repeated measurements; E = end of daily measurements; I = individual interviews with adolescent and ST therapist; T0 = pretreatment assessment; T1 = posttreatment assessment; T2 = follow-up assessment; wk = week

risk of suicide requiring immediate intervention, (3) trauma from serious violence, (4) epilepsy or serious problems regarding motion sickness, and (5) insufficient understanding of the spoken and written Dutch language.

All ST therapists were extensively informed about the study procedures and the in- and exclusion criteria for participants. When an adolescent was referred to ST and was thought to be eligible to participate in the study based on pre-screening, ST therapists informed the adolescent about the study. In case the adolescent was interested and gave oral permission to be approached, an informed consent appointment was planned. Written informed consent was signed when the adolescent agreed to participate in the study. In the case of a minor (12–16 years), parents, or a legal guardian, had to sign a written informed consent as well. Adolescents who did not agree to participate in the study, did not start with ST and received care as usual.

Setting

The participants were recruited across two different settings. The first setting entailed a rural (secured) residential youth care facility in the province of South Holland, the Netherlands (residential treatment setting). This location also included a school providing secondary special education, partly connected to the residential care facilities. Participants were recruited from both the (secured) living groups as well as from the school. The other setting entailed a specific department of a rural youth care provider in the province of North Holland, the Netherlands (ambulatory treatment setting). This department provides resocialization trajectories for adolescents with severe behavioral problems who dropped out of school, for example. The populations of both settings are characterized by multiple and complex externalizing behavioral problems, sometimes in combination with internalizing problems. Psychiatric and addiction problems can exist as well. The adolescents often come from unstable and problematic family situations and delinquent behavior can be present.

Participant Characteristics

Within the first setting (residential treatment), 14 adolescents have been screened between September 2021 and October 2023. Of these 14 adolescents, 7 were not approached for study participation due to a negative screening outcome, or because the organization decided that execution of the study was no longer feasible for their organization. Of the other 7 adolescents that were approached, 4 adolescents signed informed consent to participate in the study, of which 3 were included and 1 was not included because parents did not agree with participation of their child.

Participant 1 dropped out during the baseline phase. The intervention could not start because of covid quarantine, and in the meantime the problems of the participant worsened making other care more appropriate. Participant 2 dropped out between inclusion and pre-measurement. He lost his motivation to participate because it took too long to obtain permission from his legal guardian. Participant 3 dropped out during the baseline phase. Her problems worsened throughout the baseline phase, leading to a transfer to another care facility. None of the 3 participants completed measurements in at least two phases, therefore they will not be further discussed.

Within the second setting (ambulatory treatment), 25 adolescents have been screened between November 2022 and December 2023. Of these 25 adolescents, 12 were not approached for study participation due to a negative screening outcome, practical reasons, or because study inclusion had terminated in the meantime. Of the other 13 adolescents that were approached, 5 adolescents signed informed consent to participate in the study, and all 5 were included.

Participant 1 dropped out during the intervention phase. He disappeared for unknown reasons and upon return he was not motivated anymore to continue participation. Participant 2 completed the entire study period. Participant 3 dropped out during the intervention phase, because his motivation declined as it was not allowed that the time spent for ST and the study would shorten his penalty. Participant 4 dropped out during the intervention phase as well because another treatment focus became more important. Participant 5 completed the entire study period. Participants 3 and 4 did not complete measurements in two phases or more, therefore they will not be further discussed. Participants 1, 2, and 5 completed measurements in at least two phases, meaning that these participants will be further discussed.

Ethics

This study was conducted in accordance with the principles of the World Medical Association Declaration of Helsinki (World Medical Association, 2013) and in accordance with the Medical Research Involving Human Subjects act. Ethics approval for this study was granted by the independent Medical Ethical Committee of Vrije Universiteit medical center (reference number: 2021.0114).

Measures

Daily Repeated Measures

The daily repeated measures were assessed by means of a 14-item questionnaire, constructed by the authors. We assessed cognitive distortions (3 items), treatment

motivation (5 items), and emotion regulation (2 items). These constructs were assessed once a day at the end of the day, randomly ordered, in the format of a digital self-report questionnaire for the adolescents. The items were based on the How I Think questionnaire (HIT; Nas et al., 2008), the Adolescent Treatment Motivation Questionnaire (ATMQ; Van der Helm et al., 2013), the Regulation of Emotion Systems Survey (De France & Hollenstein, 2017; Medland et al., 2020), the Ecological Momentary Assessment (EMA) item repository (Kirtley et al., 2019) and the Difficulties in Emotion Regulation Scale (DERS; Gratz & Roemer, 2004). The questionnaire was administered through the m-Path app (Mestdagh et al., 2023), which was installed on the mobile phone of the respective adolescent during pre-measurement. Data resulting from the m-Path app was transferred to a Castor Electronic Data Capture (Castor EDC) database (van Linschoten, 2023).

Responses to each item were recorded using a visual analogue scale (VAS). Unlike a Likert scale, which offers fixed response options, a VAS typically consists of a horizontal line representing a continuum between minimum and maximum outcomes. Respondents mark the point that reflects their perceived position along this continuum. The VAS has been shown to be a reliable, valid, and easy-to-administer measure across several research fields (Brazier & Ratcliffe, 2017; Miller et al., 2003). We chose the VAS over the Likert scales from the original questionnaires to offer participants greater flexibility in answering the daily questions. For statistical purposes, the scale ranged from 0 (minimum) to 100 (maximum). The number on the scale was not visible to participants. While the reliability and validity of adapting the respective items from a Likert scale to a VAS in our study have not yet been fully established, this approach builds on evidence supporting the robustness of VAS measures in similar contexts (Arts et al., 2024; Rhee et al., 2017).

Other items of the daily questionnaire assessed how a participant was feeling, and whether they had a ST session that day. The questionnaire ended with an open question so participants could communicate something else if they wanted. The questionnaire was checked with two adolescents from one of the study sites (who would not participate in the SCED study), to receive feedback on the time needed to fill in the questionnaire, the questions, and the idea of daily assessment. Based on the feedback received, no alterations were made. In Supplementary Material 1 we elaborate on the construction process of the questionnaire. The complete daily questionnaire, including the response scales and original questions if applicable, can be requested from the authors.

Pre-, Post-, and Follow-up Measures

During pre-, post-, and follow-up assessment, the following constructs were assessed: cognitive distortions, treatment motivation, mentalization (operationalized as reflective functioning) and perspective-taking. We used the HIT (Nas et al., 2008), the ATMQ (Van der Helm et al., 2013), the Reflective Functioning Questionnaire for Youths (RFQY; Ha et al., 2013), the Self-Reflection and Insight Scale for Youth (SRIS-Y; Sauter et al., 2010) and the Perspective Taking (PT) subscale of the Interpersonal Reactivity Index (IRI; Davis, 1980, 1983).

The measurements were assessed through direct data-entry in a Castor EDC database (van Linschoten, 2023). Additionally, after finishing the intervention phase, interviews with adolescents and ST therapists were conducted to explore the overall experiences regarding ST and VR. The interviews were recorded using professional recording equipment. Personal communication throughout each participant's study period, both with each participant and the ST therapist, served to gather additional information. A comprehensive description of the instruments used can be found in Supplementary Material 1.

Intervention

ST consisted of seven sessions that were divided over two modules. Each session lasted between 45 and 60 min and was executed individually. The first session served as an intake, the last session as a closing session. In module 1, several perspective-taking exercises were done, which were based on a 360° VR video (see Fig. 2 for a screen capture). By means of this video, youth were presented with a fictional situation depicting (violent) disruptive behavior. A detailed description of the VR video was presented in our recently published paper (Klein Schaarsberg et al., 2024). Several characters played a role in the story and each of the three sessions centered around one of the characters. The exercises were guided by the therapists, whilst providing room for adolescents to take charge by letting them choose different aspects such as which character is discussed when.

Individual backstories were created for each character, using several building block cards that focused on, for example, family, school, and living situation. These cards presented different questions to help adolescents to create a personal story behind the chosen character, such as "What kind of neighborhood does [character] live in?", "What does [character] spend the most money on?", "Is [character] a family person?", or "With whom does [character] call or text the most?". The therapists let the adolescents choose different cards and then they guided the adolescents in creating the backstory by asking the questions and providing

Fig. 2 Screen capture VR video ST

further prompts. When the backstory had been created, adolescents took on the perspective of the respective character and reflected on their experience of the situation. They were explicitly encouraged to connect visible behavior to inner mental states such as thoughts and emotions, emerging from the created backstory. Following, adolescents were asked to reflect on similarities and differences between the character and themselves regarding the provided situation.

In module 2, adolescents chose a personal experience as a basis for the same perspective-taking exercises. They visualized their experience using a VR street view app (Wander, Parkline Interactive, 2019). This app visually transported adolescents to the exact location where their experience had occurred. Using this visual cue, the adolescents described the situation aloud. While they were immersed in the visualization, the VR glasses streamed to an additional screen, allowing the therapist to observe in real-time. The therapist could also ask questions to support the visualization, for example. For further details regarding ST, we refer to the protocol and the description of ST's development (Klein Schaarsberg et al., 2022, 2023). For additional information about the equipment used and how procedural fidelity was monitored, please see Supplementary Material 1.

Analysis

All data was summarized at individual participant level. For the daily repeated measurements, data was visualized graphically using the shiny app Single Case Design website (Bouwmeester, 2021) and inspected accordingly. For this inspection, we deliberately chose a descriptive approach rather than to follow available SCED guidelines for visual analysis such as reported by Lane and Gast (2014) including quantification of change in trend, level, and stability of data assessed within and between conditions. Both the nature and conduct of our study did not lend themselves adequately to the application of these guidelines, as the data contained many missing values, intrapersonal outliers, and fluctuations. Therefore, we decided to describe only the visible patterns and limit quantification of the data. Qualitative outcomes were approached alike and described per participant, to illustrate experiences and processes.

Data collected at pre-, post- and follow-up assessment were analyzed by means of a Reliable Change Index (RCI; Jacobson & Truax, 1991). When the RCI exceeded (-)1.96 it was assumed that a statistically reliable change had occurred between two phases, using a 0.05 significance level (Jacobson & Truax, 1991). Because the HIT provided a standardization that classifies scores as non-clinical, borderline clinical or clinical, we chose to use this standardization to indicate clinical significance rather than to focus on an RCI. No clear norms were available for the remaining questionnaires, so we calculated RCIs for those outcomes. A detailed description of how the RCIs were calculated can be found in Supplementary Material 1.

Results

To avoid information overload in the paper itself, we have chosen to provide a summary of the results of the three participants in the main text. We will highlight a number of results per participant, complemented by an overarching synthesis of the results found across the three participants. Elaborative descriptions about participants' backgrounds and their study trajectories are given in Supplementary Material 2. Readers interested in a comprehensive exposition of all results can find so in Supplementary Materials 3

Table 1 Participant demographic information

Participant	Jason	Steve	Aron
Sex	Male	Male	Male
Age	16	16	17
Country of birth	The Netherlands	The Netherlands	The Netherlands
Country of birth mother	African country	South American country	European country
Country of birth father	African country	Unknown	Unknown
First language	Dutch	Dutch	Dutch
Highest education completed	Vocational education	Primary education	Primary education
Current education	Secondary vocational education, level 1 of 4	Vocational education	Secondary vocational education, level 2 of 4
Living situation	Living at home, with mother and younger sister.	Living at home, with mother and older sister.	Living out-of-home, at a care facility in Amsterdam.
Parental marital status	Divorced	Divorced	Unmarried
Contact with the police and justice system	Yes	Yes	Yes

(Jason), 4 (Steve) and 5 (Aron). The names that are used to describe the participants are fictitious. Demographic information can be found below, in Table 1.

Jason

Jason was a 16-year-old boy, included in the study on January 10, 2023. Jason was proposed to participate in ST because of violent delinquent behavior, both as victim (for which he started trauma therapy) and as offender. There were suspicions of cognitive distortions that might have played a role in his delinquent behavior, particularly in his last offence. He appeared to be motivated for treatment.

Daily Measurements

In total, Jason participated in the study for 74 days, from the first completed measurement until the last (21 days baseline, 53 days intervention, 0 days follow-up). The ST sessions were executed on day 22 and 30. Jason completed 22 daily questionnaires, of which 14 during baseline (7 missings) and 8 during intervention (45 missings).

Regarding cognitive distortions it was expected that scores would decrease throughout the study period. The baseline phase of the three items showed more variation in scores than those in the intervention phase, which followed a more stable course. Item 4 (Fig. 3), assessing whether Jason perceived other people to try and purposefully bother him that day, showed the most favorable trend. Although the lowest scores from the baseline did not recur during the intervention, the high peaks did not recur either. Therefore, comparing baseline to intervention, it seemed that Jason both did and did not interpret other people as trying to get in his way during baseline, whereas he less interpreted other people as such during the intervention phase.

Considering emotion regulation, illustrated by items 10 and 11 (Figs. 4 and 5), we expected scores to increase for item 10 and to decrease for item 11. Both items showed a positive trend, with a more stable pattern during the intervention phase. For item 10, the lowest scores did not return in the intervention phase, but neither did the highest scores. The scores indicated that, during the intervention phase, Jason more often looked at things from a different perspective than during the baseline phase. For item 11, it appeared

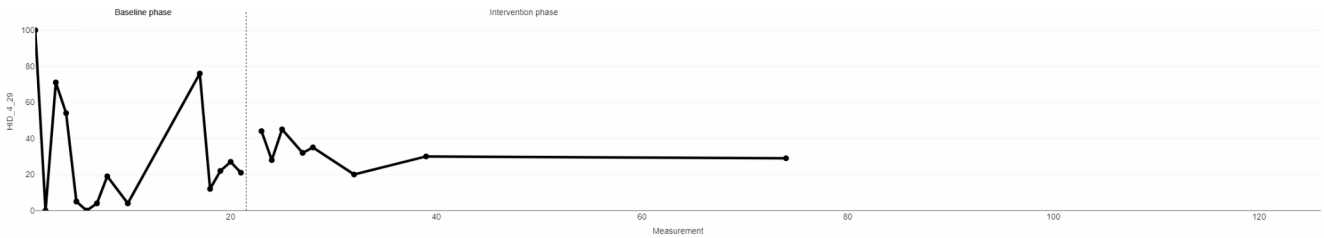

Fig. 3 Item 4. People tried to bother me today – totally disagree (0) → totally agree (100)

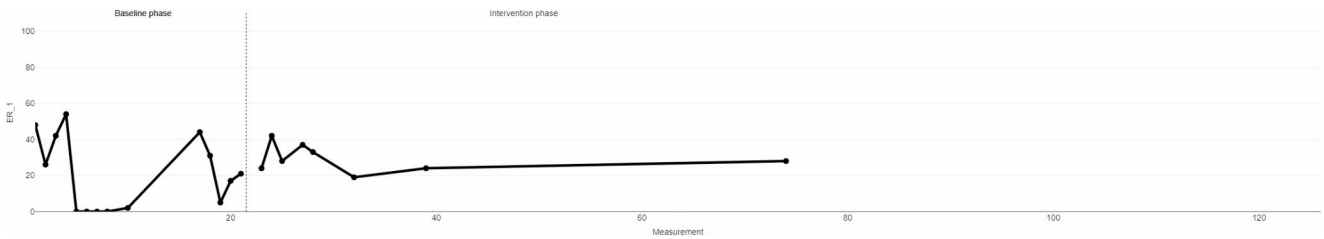

Fig. 4 Item 10. In response to my emotions today, I looked at things from a different angle – not true (0) → true (100)

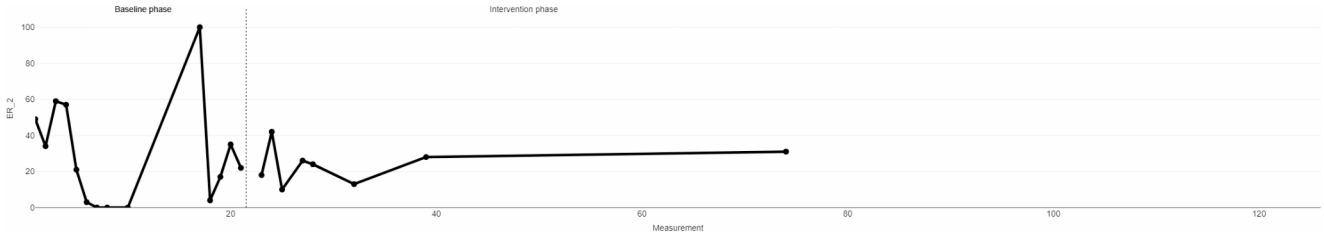

Fig. 5 Item 11. When I am upset, I lose control over my behavior – not true (0) → true (100)

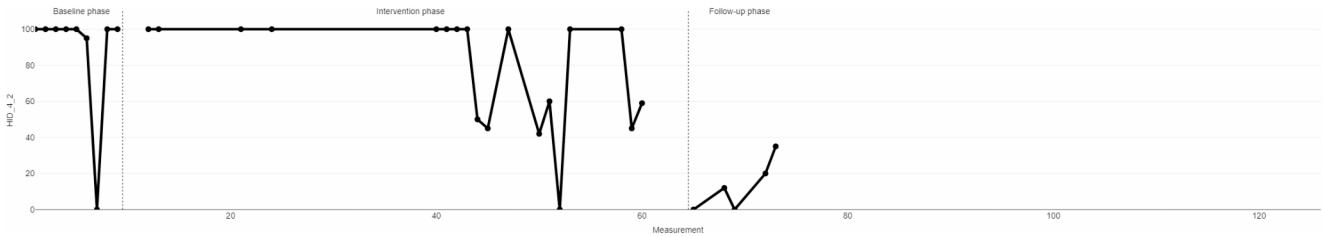

Fig. 6 Item 2. If I lose my temper, there is nothing I can do about it – totally disagree (0) → totally agree (100)

Jason perceived to have more control over his behavior when he would get upset throughout the intervention phase, also towards dropping out. Again, for a comprehensive overview of the outcomes, please see Supplementary Material 3.

Pre-, Post-, and Follow-up Assessment

For Jason, only pre-measurement outcomes were available meaning no comparison could be made. Regarding the HIT, most scores could be classified within the borderline clinical or clinical range, indicating serious problems on several domains.

On the ATMQ, Jason scored above the 75th percentile at baseline, with a score of 2,82, indicating high motivation. His total score on the RFQY was 8,74, out of a maximum score of 12. Regarding the SRIS-Y, Jason's score on the Self-Reflection skills subscale was 52 out of 66, and his score on the Insight subscale was 25 out of 36. The score of 14 on the perspective-taking scale of the IRI indicates below average perspective-taking skills.

Qualitative Data

Because Jason dropped out of the study during the intervention phase, no interview could be conducted. Therefore, no qualitative data were available.

Steve

Steve was a 16-year-old boy, included on February 13, 2023. Steve was proposed to participate in ST because of substantial problems with perception and interpretation, related to his violent behavior. He appeared to be motivated for participation.

Daily Measurements

In total, Steve participated in the study for 73 days, from the first completed measurement until the last day of the follow-up phase (9 days baseline, 55 days intervention, 9 days follow-up). The ST sessions were executed on day 10, 24, 30, 37, 45, 52, and 64. Steve completed 32 daily questionnaires, of which 9 during baseline (0 missings), 18 during intervention (37 missings), and 5 during follow-up (4 missings).

For cognitive distortions, it was expected that scores would decrease. Especially item 2 (Fig. 6) showed an interesting pattern. Apart from one outlier in the baseline phase, the scores were all as high as possible throughout the first half of the intervention phase, indicating that Steve totally agreed with the statement. Then, the scores started to fluctuate, exactly when Steve transitioned from module 1 to 2 and thus started to discuss his personal experience; the last session of module 1 was on day 38, the first of module 2 on 45. Steve started to agree less with the statement that there is nothing he can do about losing his temper, and this trend continued in the follow-up phase. In this last phase no high scores were reported anymore. Regarding item 3 (Fig. 7) the intervention phase only showed totally disagreeing scores. This was a positive contrast with the baseline and follow-up phases, which did show agreeing scores.

The outcomes for the two emotion regulation items showed considerable spread and fluctuation. We expected scores to increase for item 10 (Fig. 8) and to decrease for item 11 (Fig. 9). For both items, a similar pattern could be observed, with deviating or fluctuating scores in the second

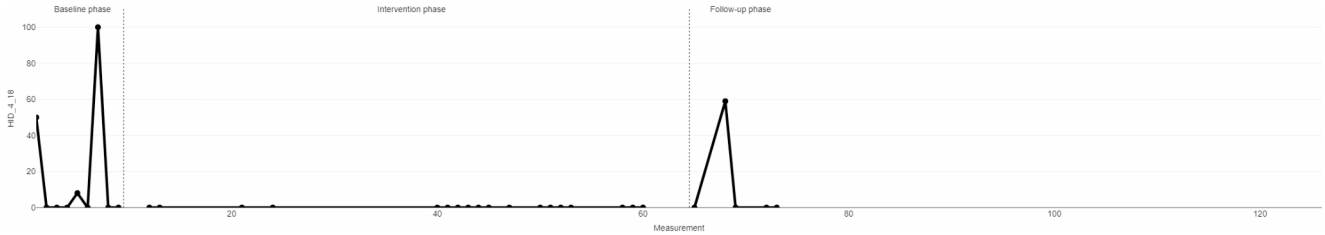

Fig. 7 Item 3. No matter how hard I try, I can't help getting in trouble today – totally disagree (0) → totally agree (100)

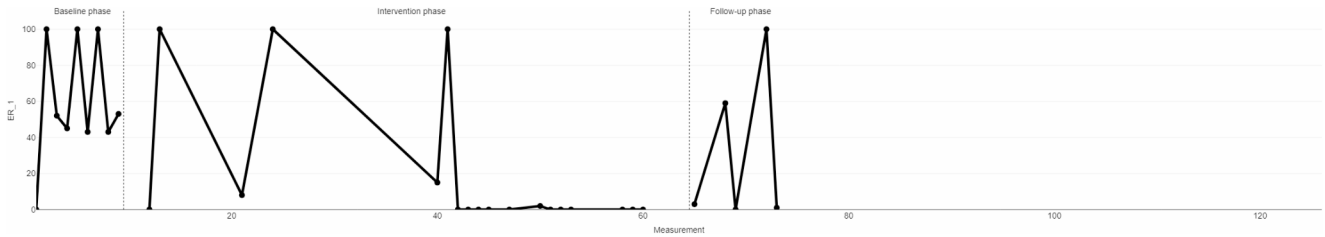

Fig. 8 Item 10. In response to my emotions today, I looked at things from a different angle – not true (0) → true (100)

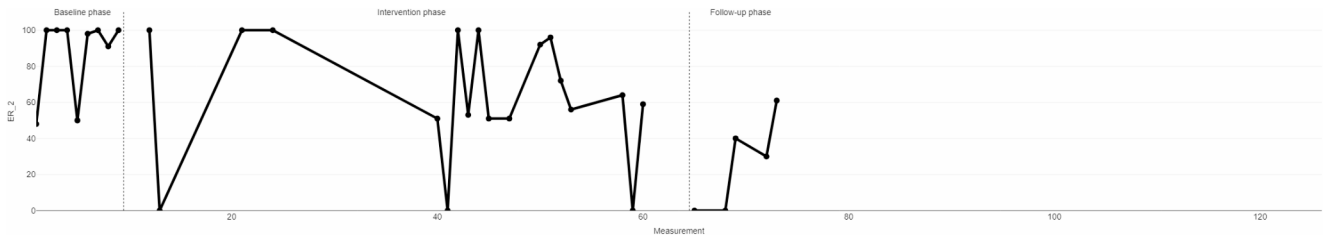

Fig. 9 Item 11. When I am upset, I lose control over my behavior – not true (0) → true (100)

half of the intervention phase when Steve's personal experience was discussed. Again, the confrontation of discussing his personal experience may have played a role in the patterns observed. Regarding item 11 specifically, Steve appeared to perceive more control over his behavior when he would get upset at this point. This seemed to be carried over into the follow-up phase. Again, for a comprehensive overview of the outcomes, please see Supplementary Material 4.

Pre-, Post-, and Follow-up Assessment

For Steve, pre-, post-, and follow-up measurements were available. The tables displaying all scores and RCIs are presented in Supplementary Material 4. The outcomes are summarized below.

Almost all HIT-scores fell within the non-clinical range, indicating little problems regarding cognitive distortions. However, for the subscale Blaming Others, the score deteriorated from pre- to post-measurement, falling in the clinical range at post-measurement. A week later the score had improved a little, falling within the borderline-clinical range. The scores for Physical Aggression and Stealing had deteriorated into the borderline clinical range at post-treatment, having improved again into the non-clinical range a week later at follow-up. The score for Lying improved from pre- to post-treatment.

The changes in Steve's scores on the ATMQ were statistically unreliable according to the computed RCIs, indicating no reliable change in motivation for behavior change. Based on the computed RCIs for Steve's scores on the RFQY no reliable change in reflective functioning could be noted as well. The changes in Steve's scores on the Self-Reflection subscale of the SRIS-Y were statistically reliable, suggesting a reliable increase in self-reflection. The computed RCIs for Steve's scores on the Insight subscale of the SRIS-Y indicated that both the increase and decrease in insight from pre- to post-measurement and from post-measurement to follow-up respectively were statistically reliable. Lastly, Steve reported increasing scores on the PT-IRI. According to the computed RCIs the increase from pre- to post-measurement was statistically reliable, suggesting an increase in perspective-taking tendencies.

Qualitative Data

Both Steve and his ST therapist participated in the interviews at post-measurement. Both mentioned they enjoyed

doing ST. Steve said that his participation has primarily helped him to stand in someone else's shoes. This matched with his increased scores on the PT-IRI. Switching perspectives was also what interested his ST therapist the most. Furthermore, Steve stated that ST has helped him to think but also talk more about his own thoughts and feelings, both during treatment but also at home:

> Because now, when I do not feel good for example, I think about whether something has happened and before, I thought 'I'm just like that', so it did help me to think about my feelings.… Yeah, I can just talk about it more, at first, I just thought 'no I don't want to talk about it', but now I just open myself up more because I notice it helps me to be more open.… That is really important.

In turn, this contributed to feeling happier, which is also seen in item 1 of the daily questionnaire. However, the ST therapist mentioned that Steve had also done Eye Movement and Desensitization and Reprocessing (EMDR) therapy before ST, and she thought that this probably largely contributed to the improvements in his mood.

The added value of VR was mainly that it gave a more realistic touch to the exercises. Both Steve and his ST therapist were impressed with the effect the street view component had, which was used to discuss a personal experience from Steve. The ST therapist explained:

> I obviously knew the reason this boy is coming to us [ambulatory treatment setting], of course I knew that situation very well, but I had never actually seen it before, and I thought that was really of added value.… He could completely explain 'yeah that's where the neighbor came from and there was this and there was that' so that was really nice.… I also almost had the idea that he was like 'oh I'm really back at that place', he also found it kind of thrilling and he also indicated that he had never been back there before.

Furthermore, directly after the first session in which they used the street view component, the ST therapist called the researcher about the session. She said that Steve, by taking on the perspective of the neighbor as a bystander, realized that he had never apologized to this bystander. He stated that perhaps he should still do so because it probably must have been very intense for this person:

> He said to me: Shouldn't we arrange for some flowers or something like that for that neighbor, or can't we do that anymore after 3 years?.

She was extremely impressed by this realization.

Regarding the VR video, it was mentioned that it could be improved by creating a more intense and streetwise scenario. The current version was a bit too mellow, in relation to what most youths are used to.

Aron

Aron was a 17-year-old boy, included on October 12, 2023. He was proposed to participate in ST because of having trouble empathizing with others and talking mainly from his own point of view. He showed difficulty assessing how things are for someone else, making the perspective changing exercises particularly relevant.

Daily Measurements

In total, Aron participated in the study for 124 days, from the first completed measurement until the last (17 days baseline, 99 days intervention, 8 days follow-up). Aron completed 20 daily questionnaires, of which 13 during baseline (4 missings), 7 during intervention (92 missings), and 0 during follow-up (8 missings).

Regarding cognitive distortions, we expected a decline in scores throughout the study period. Items 2 (Figs. 10) and 4 (Fig. 11) showed some variability in the beginning, and already during the baseline phase scores declined. They stayed low during the intervention phase, apart from item 4's last measurement. The scores indicated that Aron increasingly thought that he could do something about losing his temper, and that he less interpreted other people as bothering him.

The scores for motivation (items 5 through 9) were expected to increase throughout the study period. One item,

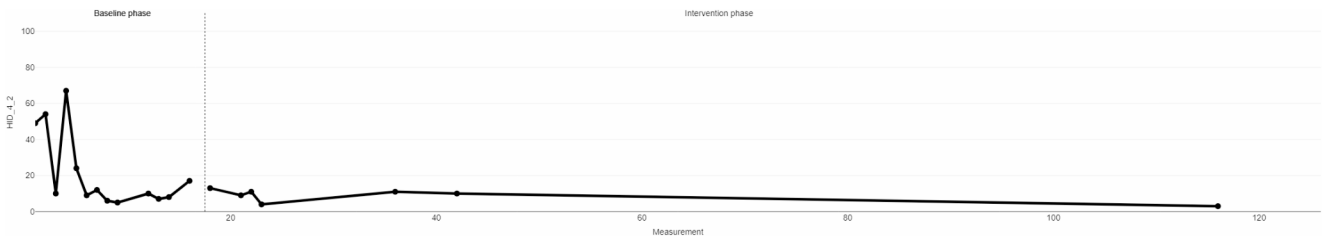

Fig. 10 Item 2. If I lose my temper, there is nothing I can do about it – totally disagree (0) → totally agree (100)

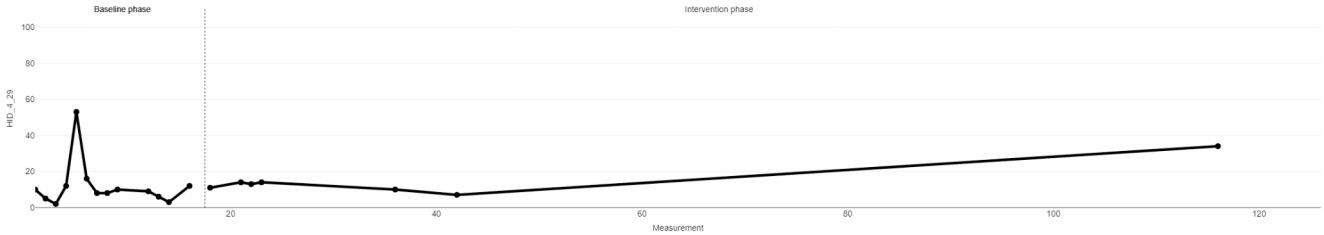

Fig. 11 Item 4. People tried to bother me today – totally disagree (0) → totally agree (100)

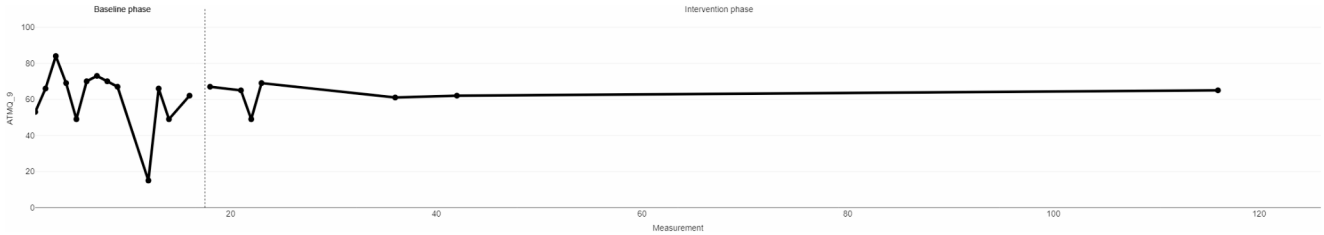

Fig. 12 Item 8. I thought about my behavior today – not true (0) → true (100)

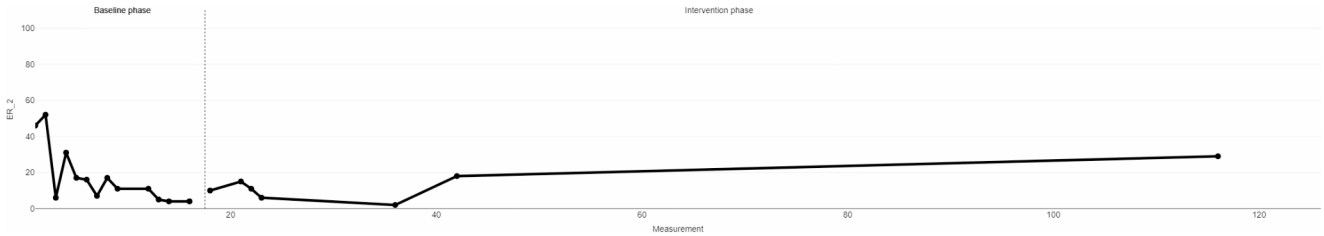

Fig. 13 Item 11. When I am upset, I lose control over my behavior – not true (0) → true (100)

number 8 (Fig. 12), showed this expected pattern, indicating that Aron started to think more about his behavior throughout his participation in ST.

Looking at emotion regulation, item 11 (Fig. 13) showed a cautiously positive pattern, suggesting Aron perceived to have more control over his behavior when he would get upset. However, it remains unclear to what extent this can be attributed to the intervention, since the change was already observable during the baseline phase. Furthermore, in the second half of the intervention phase scores started to slightly deteriorate again. Again, for a comprehensive overview of the outcomes, please see Supplementary Material 5.

Pre-, Post-, and Follow-up Assessment

For Aron, pre-, post-, and follow-up measurements were available. The tables displaying all scores and RCIs are presented in Supplementary Material 5. The outcomes are summarized below.

All HIT-scores fell within the non-clinical range, apart from the score for Assuming the worst (borderline clinical range), indicating little to no problems regarding cognitive distortions. At post-measurement and follow-up all scores fell within the non-clinical range, indicating no problems.

Aron's scores on the ATMQ increased from pre-measurement to follow-up and the RCI indicated that this increase was statistically reliable. This suggests a reliable increase in motivation for behavior change. Based on the computed RCIs for Aron's scores on the RFQY no reliable change in reflective functioning could be noted. The changes in Aron's scores on the Self-Reflection subscale of the SRIS-Y were statistically reliable, suggesting a reliable decrease in self-reflection. The computed RCIs for Aron's scores on the Insight subscale of the SRIS-Y indicated that the increase in insight from pre- to post-measurement and from post-measurement to follow-up was statistically reliable. Lastly, Aron reported a decrease in perspective-taking tendencies from pre- to post-measurement, which was statistically reliable based on the RCI. His score increased from post-measurement to follow-up, but this change was not statistically reliable.

Qualitative Data

Both Aron and his ST therapist participated in the post-treatment interviews. Aron stated that he only participated because his court-imposed conditions mandated him to come to the treatment facility. He was fine with filling this

time with participating in ST, but he said he did not learn anything from it, and he did not notice anything to be different. The only thing that was changed, was the new routine of completing the daily questionnaires. Simply put, he said he did not care that much about it all. However, he also mentioned that he found the perspective-taking exercises quite interesting. Aron's ST therapist explained that she did recognize this attitude from him. Despite Aron himself not being very disclosing, she could tell that he really enjoyed it. Especially the last sessions, which centered around his personal experiences:

> Aron was the same way with me at first, he is not very talkative with strangers at all… when we discussed it, he did say he just really liked it. Because now he could also make it visual, that which he has always told me over the past six months during which we talked about, for example, his offence. Or his neighborhood and what he did there, so he thought it was really fun to show that too using Wander.

Moreover, an important added value of using street view in Aron's case was that he was able to use the application to visualize a situation in a way that would never have been possible in real life; several location bans prohibited him from going to these places. This could not have been replaced with looking at Google Maps on a smartphone for example, according to Aron's ST therapist:

> I think because he could also see that I could watch along and he could just look with his head where he wanted to look, it is all very fast. But I can also follow that speed right away because I can watch everything live along. So, I think street view on a laptop or on a mobile would have been too difficult.

Concerning the other VR element, the video, Aron only stated that it should be improved, without giving specific ideas about relevant improvements. His ST therapist, on the other hand, explained that she primarily thought the video was not intense enough. Nowadays, violence amongst youths simply not only entails fists, weapons really do also play a role, even if they are only used to threaten.

Lastly, social desirability should be considered in Aron's case. It is striking, for example, that at the follow-up measurement every subscale score of the HIT questionnaire is the same, coming from the fact that Aron only gave the same answer to all relevant questions, without any nuance. His ST therapist also stated:

> I think he can be very socially desirable and knows when to be so and with whom to be so.

This means that, even though we established statistically reliable changes on different outcomes, and the HIT questionnaires showed little to no problems for example, the question arises to what extent these changes and outcomes reflect Aron's reality.

Overarching Synthesis

Considering the outcomes of the daily measurements of the three participants collectively, there appeared to be positive changes in cognitive distortions and emotion regulation. However, these findings should be interpreted cautiously as they may be influenced by the study's limitations. Nonetheless, looking at the items from both constructs, at least one item from each participant showed a positive pattern. Other items showed either no change, or a slightly negative change in pattern. Furthermore, cautiously positive patterns for 2 of the 3 participants in terms of their mood were noted (Jason and Steve). Considering motivation, we could not distill a positive effect of participating in ST. We observed either no change in scores, or a negative change, apart from 1 item from 1 of the participants (Aron).

The pre-, post-, and follow-up assessments of Steve and Aron showed different outcomes. Regarding cognitive distortions, almost all HIT-scores fell within the non-clinical range, indicating little problems. For Steve specifically, some scores had deteriorated at post-assessment, which had improved again at follow-up. On motivation, Steve did not show a statistically reliable change on the ATMQ, whereas Aron showed a statistically reliable increase. This increase was not consistent with his daily measurements. Looking at the RFQY, both participants did not show a statistically reliable change in reflective functioning. Regarding the SRIS-Y, Steve showed a statistically reliable increase in self-reflection. For insight, a statistically reliable increase could be noted from pre to post assessment, followed by a statistically reliable decrease in insight from post to follow-up assessment. Aron showed opposing results, with a statistically reliable decrease on the Self-Reflection subscale and a statistically reliable increase on the Insight subscale. The same was observed for perspective-taking; Steve's scores on the PT-IRI reflected statistically reliable improvement, whereas Aron's score indicated a statistically reliable decrease in perspective-taking skills.

Technical issues pertaining to the VR component emerged across the three participants. The main problem was failing attempts of streaming the VR footage to another screen, so a trainer could watch along with a participant. Another issue was stalling of the video when playing it, for example. These issues could not always be fixed on the spot,

hindering execution as planned. Therefore, the technological constraints need further development to improve future execution.

Discussion

The aim of the present study was to explore the added value of ST to forensic youth care. The primary focus was on the use of mentalization as a main treatment mechanism, and how such an approach could help to address cognitive distortions and motivation for behavior change as treatment responsivity indicators. In total, 8 participants were included in the single-case study, of whom 5 participants dropped out. Three participants completed sufficient data for visual inspection of the results. Two of them completed the entire study period. Considering the outcomes of the 3 participants included in the analysis, all 3 trajectories showed different patterns, both in terms of daily assessment and the pre-, post-, and follow-up assessments. Besides these different patterns, the daily measurements of all participants showed at least one item on cognitive distortions and emotion regulation that changed positively throughout the study period. However, on motivation, no positive effect of participating in ST was found. The added value of VR to ST was clearly stated in the interviews, although areas for improvement were unanimously mentioned.

The positive changes that seemed visible on cognitive distortions and emotion regulation could indicate that, for the adolescents who currently participated, using a relatively short and low-threshold training program such as ST may be beneficial to their treatment responsivity, and potentially to certain treatment outcomes as well. This is hopeful, considering the complexities that are involved with treatment in forensic youth care, though the results warrant careful interpretation given the exploratory nature of the study. To some extent, these results are in line with a recent study, in which reductions in hostile attribution bias were found following a short 5-session modification training (Van Bockstaele et al., 2020). Hostile attribution bias has been widely studied as a specific biased way of attributing meaning to social situations and behavior, and can be considered to be an example of cognitive distortions (Barriga et al., 2008; Brugman et al., 2011; Verhoef et al., 2019). It is likely that, through reappraising hostile inferences and replacing them with other, less hostile interpretations, behavioral reactions can become more controlled and less hostile as well (Gagnon & Rochat, 2017).

Less favorable outcomes regarding motivation were found. Positive effects on motivation, at least concerning daily assessment, were little to not observed. In some cases, certain motivation items, such as trusting therapists or the perceived benefit of going to the treatment facility, already started off on a positive note or even showed a ceiling effect. Currently, it is not exactly clear what these positive observations can tell us. Considering the trajectories of Steve and Aron, who completed the intervention, and the fact that they reported their guidance as being useful in combination with reporting at least some trust in their therapists, it may be the case that some form of a trusting relationship between an adolescent and their therapist is a precondition in order to execute ST. However, Jason's results are not in line with this possibility. He did not complete the intervention but did report to find his guidance useful as well. Moreover, he reported higher scores regarding trusting his therapist than Steve and Aron. Therefore, both the role of as the effects on trust in relation to ST and motivation, for example, remain somewhat unclear. Furthermore, because of the mentalization-based approach, the primary aim of ST was to challenge youngsters to interpret behavior of both themselves and others, by reflecting on underlying factors such as thoughts and emotions. As a result, the wording of the items on motivation that directly or indirectly related to this aim may have been too abstract. The words "behavior" and "problems" in questions such as "I want to change my behavior" and "I talk about my problems" can refer to different behavior and problems. This may have made it more difficult to specify a particular motivation for behavior change and therefore may have complicated answering the questions. Moreover, it is possible that these generally formulated questions led to a more trait-like rather than state-like assessment, limiting meaningful outcomes on a daily basis (Myin-Germeys & Kuppens, 2022).

A common pattern that was seen in the daily measurements is that the variability in the results seemed to decrease as time passed. It is not exactly clear what this indicates. It could be that these patterns are the result of a time-bound pattern seen more often in repeated assessment, especially in repeated subjective assessment. This pattern is known as the initial elevation bias, in which consecutive subjective reports may tend to exhibit higher values during initial assessments. These higher values are followed by a slight decline that eventually stabilizes (Myin-Germeys & Kuppens, 2022; Shrout et al., 2018). It is unclear why this bias occurs, but in any case, it means we should interpret the current findings with express caution.

The initial elevation bias may also play a part in the fact that common SCED guidelines recommend researchers to await baseline stability, before transitioning into the consecutive phase (Kazdin, 2011; Kratochwill et al., 2010). In our study, transitioning from baseline to intervention did not depend on baseline stability. In addition, we did observe some changes in the daily measurements already occurred during baseline (possibly due to an initial elevation bias,

as previously explained). As a result, we currently cannot make convincing statements about whether the visible patterns could actually be attributed to participation in ST specifically. The question is, however, to what extent awaiting baseline stability would have reduced this problem. Recently, this recommendation has been questioned due to a lack of empirical evidence to support this recommendation (Lanovaz & Primiani, 2023). Moreover, we should consider that certain concepts fluctuate over time, although, in theory, there are assumed to be static (Myin-Germeys & Kuppens, 2022). Therefore, it seems reasonable to consider the actual relevance of baseline stability in psychologically oriented studies. Perhaps it is of greater importance to obtain a representative snapshot during baseline, which may include realistic fluctuations.

Considering the VR component of ST, being the 360° video and the VR street view application, the qualitative data showed that these relatively simple VR tools added a lot of value to the exercises and yielded promising results. In particular, the video and street view application seemed to foster the conversation between the therapist and the adolescent. Furthermore, it was of specific value for therapists to actually *see* a certain place that an adolescent was talking about, when explaining a situation that had happened there. Even if the therapists already knew the story. Moreover, in one case the use of VR street view even led to a breakthrough in the way he considered the perspective of one of the bystanders. Previous research has shown similar outcomes in terms of how a VR tool may help to facilitate therapeutic communication (Falconer et al., 2019). Findings like these are important, considering the difficulties that can emerge when pursuing adolescents to engage in therapeutic conversations (O'Keeffe et al., 2018; Stige et al., 2021).

More specifically, these qualitative results hold promise for the mentalizing approach that is central to ST. By instigating conversations about adolescents' personal lives as well as the lives of others, ample options emerge to challenge them to mentalize. Similarly, other, more interactive forms of VR exercises have been found to generate positive outcomes regarding perspective-taking (Chen et al., 2021; van Loon et al., 2018), a crucial element of mentalization (Derks et al., 2019). Considering the relatively simple forms of VR that were used in ST, not only technologically advanced and highly interactive VR tools are worth considering. In this context, the rapid developments that are ongoing in VR technology provide additional ground for curiosity and positivity toward the future. Who knows what else could be accomplished in improving care for the youngsters involved, given the more advanced possibilities that may yet be explored.

In addition to the positive and facilitating aspects of working with VR, several areas for improvement were highlighted in the interviews, including the technical part of VR. Therapists mentioned that they struggled often to connect the VR headset to a second device for streaming purposes, for example. Therapists were not always able to solve these issues by themselves, having to rely on the executive researcher for assistance. Even with technical support, it was not always managed to solve a problem, which led to frustration, for instance. Moreover, this appealed greatly to the patience of both youth and therapists, something that cannot always be mustered. Technical issues have previously been identified as a weakness of working with VR (Nijman et al., 2020). However, in other recent studies looking into VR-based interventions no to only minor technical issues have been reported (Alsem et al., 2021; Arts et al., 2023). Because the technical reliability within ST should be emphasized in the future, it may be beneficial to learn from success stories like these.

A crucial aspect to discuss in relation to our study, is inclusion and dropout. To some extent, inclusion and dropout are well-known challenges in research (Groeneveld et al., 2009). However, we aimed to include a total of 18 adolescents. Ultimately, only two participants completed the entire study period. Moreover, although we still adhered to a minimum of 3, preferably 5, data points per phase as stated by SCED guidelines (Krasny-Pacini & Evans, 2018; Kratochwill et al., 2013), we cannot ignore the high number of missing values. This means that we must be careful when interpreting the results. It also means that generalization of the current results is not possible due to a lack of a sufficient number of direct and systematic replications (Walker & Carr, 2021). Therefore, the current results should only be interpreted as initial effects, related to the specific youth who participated in this study.

An explanation could be that we did not collaborate close enough with adolescents and practitioners while designing the current study. Consequently, our ideas potentially did not fit well enough with the challenges encountered in daily practice. This could apply both to the feasibility of ST as an intervention, as to the feasibility of the study that was designed and conducted under the current circumstances. That is, it could be that the study itself has been a weak link rather than ST as an intervention. Many adolescents "dropped out" before they could even be included in the study. This suggests that aspects related to the preparatory phase of participating in research (e.g., parents or legal guardians of minors were hard to reach for informed consent purposes), as well as the study design itself (e.g., being asked to fill out a daily questionnaire), potentially did not contribute positively to the inclusion of participants. Therefore, it may be premature to conclude that implementing ST in this setting, with these adolescents, is not feasible. We believe further research, possibly with a refined protocol

based on the lessons learned from this initial study, is warranted. At the same time, it is important to critically assess the realistic expectations we might have for future research in this area.

Certain practical and logistic issues have been at play as well in relation to dropout. For example, adolescents suddenly left the facility to go to school or because they got arrested (something that was not intended to be influenced by ST), therapists' schedules abruptly changed due to an emergency or they could not leave the group in order to execute ST with an individual adolescent, or the challenges within the organization itself were such that no time could be devoted to other things. Because these issues were, at least to some extent, inherent to the current research settings, it is important to consider whether these settings are the most fruitful for certain research purposes. In the current study we were able to explore the added value of ST, including VR, within this complex context. However, for certain future research and implementation purposes, such as studying the effect of the VR components specifically, a school context would be worth considering as well, for example (Cummings et al., 2022; Richter et al., 2022).

In any case, we must acknowledge that the problems concerning inclusion and dropout may have created a selection bias. On the other hand, it seems too early to draw any definitive conclusions on whether ST as an intervention is not feasible or acceptable, or that the current study was insufficiently feasible and acceptable to conduct under the current conditions.

Future Directions

Based on the current study and its outcomes, several implications for future directions can be noted. First, ST, as a VR-based add-on intervention, appears to offer relatively simple options to incorporate mentalization into treatment and in that way potentially help to enhance the treatment responsivity of adolescents. To further enhance outcomes, it may be worthwhile to consider other forms of delivering the program. For example, in line with the significant positive effects seen after brief intensive trauma-focused treatment (De Jongh et al., 2020; Voorendonk et al., 2020), it may be interesting to implement a brief intensive ST program that is delivered within 1 to 2 weeks. Moreover, we can note positive outcomes, even when the program is not completed in its entirety. Therefore, but possibly also in a brief intensive form, ST may be of specific value in the context of the often relatively short placements within the judicial, forensic setting (Buysse et al., 2021). Nevertheless, looking at the practical constraints we encountered, options like these would require considerable organizational efforts.

Second, the addition of VR components to treatment within the context of forensic youth care seems to hold promise and therefore deserves further attention. Using VR, especially the street view component, may hold value in other forms of usage as well. For example, it may be relevant for crime analysis, during which somebody gives a factual description of the events that led to a committed crime, as well as one's thoughts, feelings, and behavior in relation to the crime. As illustrated by our study, the visual support of using VR street view may add to a reflective exercise like this. Also, VR components like the once discussed may be helpful in preparation of confrontations with possible victims, which is gaining interest by means of restorative justice practices (Rossner & Taylor, 2024; Suzuki, 2023). Considering VR, however, it is important to gain a better understanding of how to develop and use VR in a way that is both ethically responsible and suitable for the adolescents aimed at and the situations they can face. This would include a more extensive evaluation of the VR materials used, also consisting of quantitative assessment methods for example. Youth involved in forensic youth care can be used to quite some violence, unfortunately, which can make a relatively mild scenario like the current one not emotionally engaging enough. Therefore, it was suggested to incorporate weapons like guns or knives, for example. However, this is not something that can be done without question; it is not yet clear where the line lies between an emotionally engaging scenario and one that is irresponsible and possibly even harmful. Violent video games, for example, have been found to negatively impact aggression, other problem behaviors, and normative beliefs about aggression (Greitemeyer, 2022; Shao & Wang, 2019; Wei et al., 2022). There is also evidence against these negative effects available, in adults that is (Kühn et al., 2019), but caution is warranted at the very least. In addition, given the boundless opportunities presented by VR and the blurring of boundaries between virtual and real experiences, there exists a moral imperative in the development of VR experiences (Bailenson, 2018; Brey, 1999; Ford, 2001). Therefore, we must reflect on the question whether we *should* create everything.

Last, for future practice-oriented research it may be particularly beneficial to collaborate with adolescents and practitioners in every step of the research process. By collectively thinking about assessment methods and piloting these methods more extensively, for example, more room can be created to incorporate the language of adolescents themselves instead of primarily relying on therapeutic or academic language. This will likely allow for more idiosyncratically oriented research methods as well. In this way, we can try to get as close as possible to the experiences of the adolescents involved and the ways they express themselves. A good example may be found in the Youth Top Problems

measure, for example, a psychometrically sound, client-guided approach to assessment (Weisz et al., 2011).

Strengths and Limitations

An important strength of this study was the ecological validity. The study was performed under daily, uncontrolled circumstances. While these conditions do create some limitations, as will be discussed below, they also make sure that the outcomes are of direct value to clinical practice. In addition, the results showed that progress and change can be difficult to translate into a set of numbers. Therefore, our study approach, including different data collection methods, moments, and informants, helped improve our understanding of the effects. Furthermore, this triangulation decreased possible deficiencies resulting from only using self-report questionnaires (Thurmond, 2001).

Despite these strengths, several limitations must be considered as well. One of the strengths, the ecological validity, simultaneously created one of the limitations. The nature of the study prohibits attribution of the observed outcomes to either ST or other specific circumstances or therapies, for example. Moreover, we observed some changes in the daily measurements already occurred during baseline (possibly due to an initial elevation bias, as explained in the beginning of the discussion). As a result, we cannot make convincing statements about whether the visible patterns could be attributed to participation in ST. Lastly, while clinical judgment was chosen to reduce the burden on adolescents and clinicians and to accommodate the population's heterogeneity, this approach limits the generalizability of the findings to other contexts or populations. Future research should consider incorporating standardized assessments to strengthen the evidence base and to enhance comparability across studies.

Conclusions

Conducting our single-case study within the context of forensic youth care proved complex, but also feasible. Moreover, considering the three different research and treatment trajectories that emerged, this individually oriented research design may hold particular potential for studying this intricate population. The positive changes that appeared to come forward could suggest that the adolescents who currently participated have learned to think more about another person's perspective over the course of ST. Additionally, they seemed to show positive change in the amount of control they perceived to have over their behavior. This may be beneficial for adolescents' treatment responsivity. Moreover, these outcomes hold preliminary promise for the use of a mentalization-based approach, such as ST, within the complex context of forensic youth care. However, these results did not seem to necessarily lead to them wanting to work more on their own behavior and problems. Therefore, based on the current outcomes, we cannot conclude that participation in ST was helpful in increasing the adolescents' motivation for behavior change. Working with VR was deemed advantageous, despite several limiting factors. Taken together, both the future development and study of ST and the VR components are warranted, despite the difficulties of taking the first steps. VR, as a tool within ST, appears to be a valuable resource that can be seen as an addition to the therapeutic toolbox. Furthermore, the positive results that have been highlighted in the current phase of development fuel our curiosity and positivity about future (technological) advancements in forensic child and youth care.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10802-025-01313-1.

Acknowledgements The authors thank all youngsters and professionals that have participated in this study. Furthermore, they give special thanks to Erik Jongman, for his valuable role within the project team. They would also like to thank the reviewers and the guest editor for their helpful comments and suggestions on the manuscript.

Author Contributions All authors contributed to the study conception and design. Material preparation, data collection and analysis were performed by Renée Klein Schaarsberg and Nicole Klinkhamer, with contribution of Samantha Bouwmeester and Levi van Dam to material preparation and analysis. The first draft of the manuscript was written by Renée Klein Schaarsberg, with assistance of Nicole Klinkhamer. All authors commented on previous versions of the manuscript. All authors read and approved the final manuscript.

Funding This study was funded by a grant from the Dutch Ministry of Health, Welfare, and Sport (in Dutch: Ministerie van Volksgezondheid, Welzijn en Sport). The funding agency had no role in the design of the study, in the collection, analyses, or interpretation of data, or in the writing of the manuscript.

Data Availability The data that support the findings of this study are not publicly available due to their containing information that could compromise the privacy of research participants.

Declarations

Ethics Approval and Consent to Participate This study was conducted according to the principles of the World Medical Association Declaration of Helsinki [38] and in accordance with the Medical Research Involving Human Subjects act. Ethics approval for this study was granted by the independent Medical Ethical Committee of Vrije Universiteit medical center (reference number: 2021.0114). Informed consent for participation in the study was obtained from participants, as well as from parents or a legal guardian in the case of minors.

Competing Interests Garage2020, where Levi van Dam is affiliated with, developed ST and an educational program for professionals to learn about the scientific and practical background of ST. The authors

declare there are no other competing interests to disclose.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

Allen, J. G., Fonagy, P., & Bateman, A. W. (2008). *Mentalizing in clinical practice*. American Psychiatric Publishing.

Alsem, S. C., van Dijk, A., Verhulp, E. E., & De Orobio, B. (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

Arsenio, W. F. (2010). Social information processing, emotions, and aggression: Conceptual and methodological contributions of the special section articles. *Journal of Abnormal Child Psychology*, *38*(5), 627–632. https://doi.org/10.1007/s10802-010-9408-z

Arts, E., De Castro, O., Luteijn, B., Elsendoorn, E., B., & Vissers, C. T. W. M. (2023). Interactive virtual reality training to improve socio-emotional functioning in adolescents with developmental Language disorders: A feasibility study. *Clinical Child Psychology and Psychiatry*, 13591045231220694. https://doi.org/10.1177/13591045231220694

Arts, E., De Castro, B. O., Luteijn, E., Elsendoorn, B., Maric, M., & Vissers, C. T. W. M. (2024). Virtual reality training to improve socio-emotional functioning in adolescents with developmental Language disorders: A multiple baseline effectiveness study. *Social Development*, *34*(1), e12784. https://doi.org/10.1111/sode.12784

Baglio, G., Blasi, V., Sangiuliano Intra, F., Castelli, I., Massaro, D., Baglio, F., Valle, A., Zanette, M., & Marchetti, A. (2016). Social competence in children with borderline intellectual functioning: Delayed development of theory of Mind across all complexity levels. *Frontiers in Psychology*, *7*, 1604. https://doi.org/10.3389/fpsyg.2016.01604

Bailenson, J. N. (2018). *Experience on demand. What virtual reality is, how it works, and what it can do*. W.W. Norton.

Barriga, A. Q., & Gibbs, J. C. (1996). Measuring cognitive distortion in antisocial youth: Development and preliminary validation of the how I think questionnaire. *Aggressive Behavior*, *22*(5), 333–343. https://doi.org/10.1002/(SICI)1098-2337(1996)22:5%3C333::AID-AB2%3E3.0.CO;2-K

Barriga, A. Q., Landau, J. R., Stinson, B. L., Liau, A., & Gibbs, J. C. (2000). Cognitive distortion and problem behaviors in adolescents. *Criminal Justice and Behavior*, *27*(1), 36–56. https://doi.org/10.1177/0093854800027001003

Barriga, A. Q., Gibbs, J. C., Potter, G. B., & Liau, A. (2001). *How I think (HIT) questionnaire manual*. Research.

Barriga, A. Q., Hawkins, M. A., & Camelia, C. R. T. (2008). Specificity of cognitive distortions to antisocial behaviours. *Criminal Behaviour and Mental Health*, *18*(2), 104–116. https://doi.org/10.1002/cbm.683

Bateman, A. W., & Fonagy, P. (2004). *Psychotherapy for borderline personality disorder: Mentalization based treatment*. Oxford University Press.

Bouwmeester, S. (2021). *Single case designs*. In (Version 2.2) https://architecta.shinyapps.io/SingleCaseDesignsv02/

Brauers, M., Kroneman, L., Otten, R., Lindauer, R., & Popma, A. (2016). Enhancing adolescents' motivation for treatment in compulsory residential care: A clinical review. *Children and Youth Services Review*, *61*, 117–125. https://doi.org/10.1016/j.childyouth.2015.12.011

Brazier, J., & Ratcliffe, J. (2017). Measurement and valuation of health for economic evaluation. In S. R. Quah (Ed.), *International Encyclopedia of Public Health (Second Edition)* (pp. 586–593). Academic Press. https://doi.org/10.1016/B978-0-12-803678-5.00457-4

Brey, P. (1999). The ethics of representation and action in virtual reality. *Ethics and Information Technology*, *1*(1), 5–14. https://doi.org/10.1023/A:1010069907461

Brugman, D., Nas, C. N., Van der Velden, F., Barriga, A. Q., Gibbs, J. C., Potter, G. B., & Liau, A. K. (2011). *Hoe Ik Denk Vragenlijst (HID) handleiding*. Boom test uitgevers.

Buysse, W., Piepers, N., Pluijm, M., & Dijk, B. v (2021). Vrijheidsbeneming op Maat. *Opzet En Haalbaarheid monitor*. DSP-groep.

Carl, L. C., Schmucker, M., & Lösel, F. (2020). Predicting attrition and engagement in the treatment of young offenders. *International Journal of Offender Therapy and Comparative Criminology*, *64*, 355–374. https://doi.org/10.1177/0306624X19877593

Chen, V. H., Chan, S. H., & Tan, Y. C. (2021). Perspective-taking in virtual reality and reduction of biases against minorities. *Multimodal Technologies and Interaction*, *5*(8).

Choi-Kain, L. W., & Gunderson, J. G. (2008). Mentalization: Ontogeny, assessment, and application in the treatment of borderline personality disorder. *American Journal of Psychiatry*, *165*, 1127–1135.

Cornet, L. J. M., & van Gelder, J. L. (2020). Virtual reality: A use case for criminal justice practice. *Psychology Crime & Law*, 1–17. https://doi.org/10.1080/1068316X.2019.1708357

Cummings, J. R., Zhang, Y. X., Wilk, A. S., & Marcus, S. C. (2022). Mental health clinic administrators' perspectives on the impact of clinic-School partnerships on youth mental health services continuity and quality. *School Mental Health*, *14*(4), 1086–1097. https://doi.org/10.1007/s12310-022-09531-4

Davis, M. H. (1980). A multidimensional approach to individual differences in empathy. *JSAS Catalog of Selected Documents in Psychology*, *10*, 85.

Davis, M. H. (1983). Measuring individual differences in empathy: Evidence for a multidimensional approach. *Journal of Personality and Social Psychology*, *44*, 113–126. https://doi.org/10.1037/0022-3514.44.1.113

De Castro, B. O., & Van Dijk, A. (2017). It's gonna end up with a fight anyway. In *The Wiley handbook of disruptive and impulse-control disorders* (pp. 237–253). https://doi.org/10.1002/9781119092254.ch15

De France, K., & Hollenstein, T. (2017). Assessing emotion regulation repertoires: The regulation of emotion systems survey. *Personality and Individual Differences*, *119*, 204–215. https://doi.org/10.1016/j.paid.2017.07.018

De Jongh, A., Groenland, G. N., Sanches, S., Bongaerts, H., Voorendonk, E. M., & Van Minnen, A. (2020). The impact of brief intensive trauma-focused treatment for PTSD on symptoms of borderline personality disorder. *European Journal of Psychotraumatology*, *11*(1), 1721142. https://doi.org/10.1080/20008198.2020.1721142

de Wit, M., Moonen, X., & Douma, J. (2023). *Richtlijn effectieve interventies LVB. Aanbevelingen voor Het Ontwikkelen, Aanpassen En Uitvoeren Van Gedragsveranderende interventies*

voor personen Met Een licht Verstandelijke beperking. Herziene versie. In. Landelijk Kenniscentrum LVB.

Deperrois, R., & Combalbert, N. (2022). Links between cognitive distortions and cognitive emotion regulation strategies in non-clinical young adulthood. *Cognitive Processing*, *23*(1), 69–77. https://doi.org/10.1007/s10339-021-01057-y

Derks, S., Van Wijngaarden, S., Wouda, M., Schuengel, C., & Sterkenburg, P. S. (2019). Effectiveness of the serious game 'you and I' in changing mentalizing abilities of adults with mild to borderline intellectual disabilities: A parallel superiority randomized controlled trial. *Trials*, *20*, 500. https://doi.org/10.1186/s13063-019-3608-9

Drieschner, K. H., Lammers, S. M. M., & van der Staak, C. P. F. (2004). Treatment motivation: An attempt for clarification of an ambiguous concept. *Clinical Psychology Review*, *23*(8), 1115–1137. https://doi.org/10.1016/j.cpr.2003.09.003

Englebrecht, C., Peterson, D., Scherer, A., & Naccarato, T. (2008). It's not my fault: Acceptance of responsibility as a component of engagement in juvenile residential treatment. *Children and Youth Services Review*, *30*, 466–484. https://doi.org/10.1016/j.childyouth.2007.11.005

Falconer, C. J., Davies, E. B., Grist, R., & Stallard, P. (2019). Innovations in practice: Avatar-based virtual reality in CAMHS talking therapy: Two exploratory case studies. *Child Adolesc Ment Health*, *24*(3), 283–287. https://doi.org/10.1111/camh.12326

Fonagy, P., Target, M., Steele, H., & Steele, M. (1998). *Reflective-Functioning manual, version 5.0, for application to adult attachment interviews*. University College London.

Ford, P. J. (2001). A further analysis of the ethics of representation in virtual reality: Multi-user environments. *Ethics and Information Technology*, *3*(2), 113–121. https://doi.org/10.1023/A:1011846009390

Gagnon, J., & Rochat, L. (2017). Relationships between hostile attribution bias, negative urgency, and reactive aggression. *Journal of Individual Differences*, *38*(4), 211–219. https://doi.org/10.1027/1614-0001/a000238

Gergeley, G., Fonagy, P., Jurist, E., & Target, M. (2002). *Affect regulation, mentalization, and the development of the self*. Other.

Gibbs, J. C., Potter, G. B., & DiBiase, A. M. (2013). Sociomoral development for behaviorally at-risk youth: Mac's group meeting. In C. Proctor & P. A. Linley (Eds.), *Research, Applications, and Interventions for Children and Adolescents: A Positive Psychology Perspective* (pp. 225–245). Springer Netherlands. https://doi.org/10.1007/978-94-007-6398-2_14

Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the difficulties in emotion regulation scale. *Journal of Psychopathology and Behavioral Assessment*, *26*, 41–54.

Greitemeyer, T. (2022). The dark and bright side of video game consumption: Effects of violent and prosocial video games. *Current Opinion in Psychology*, *46*, 101326. https://doi.org/10.1016/j.copsyc.2022.101326

Groeneveld, I. F., Proper, K. I., van der Beek, A. J., Hildebrandt, V. H., & van Mechelen, W. (2009). Factors associated with non-participation and drop-out in a lifestyle intervention for workers with an elevated risk of cardiovascular disease. *International Journal of Behavioral Nutrition and Physical Activity*, *6*(1), 80. https://doi.org/10.1186/1479-5868-6-80

Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *2*(3), 271–299. https://doi.org/10.1037/1089-2680.2.3.271

Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, *26*(1), 1–26. https://doi.org/10.1080/1047840X.2014.940781

Ha, C., Sharp, C., Ensink, K., Fonagy, P., & Cirino, P. (2013). The measurement of reflective function in adolescents with and without borderline traits. *Journal of Adolescence*, *36*, 1215–1223. https://doi.org/10.1016/j.adolescence.2013.09.008

Happé, F. G. E. (1994). An advanced test of theory of Mind: Understanding of story characters' thoughts and feelings by able autistic, mentally handicapped, and normal children and adults. *Journal of Autism and Developmental Disorders*, *24*, 129–154. https://doi.org/10.1007/BF02172093

Harder, A. T. (2018). Residential care and cure: Achieving enduring behavior change with youth by using a self-determination, common factors and motivational interviewing approach. *Residential Treatment for Children and Youth*, *35*, 317–335. https://doi.org/10.1080/0886571X.2018.1460006

Harder, A. T., Knorth, E. J., & Kalverboer, M. E. (2012). Securing the downside up: Client and care factors associated with outcomes of secure residential youth care. *Child & Youth Care Forum*, *41*(3), 259–276. https://doi.org/10.1007/s10566-011-9159-1

Hillege, S. L., Brand, E. F. J. M., Mulder, E. A., Vermeiren, R. R., J., M., & Van Domburgh, L. (2017). Serious juvenile offenders: Classification into subgroups based on static and dynamic characteristics. *Child and Adolescent Psychiatry and Mental Health*, *11*, 67. https://doi.org/10.1186/s13034-017-0201-4

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Journal of Consulting and Clinical Psychology*, *59*(1), 12–19. https://doi.org/10.1037//0022-006x.59.1.12

Jørgensen, M. S., Bo, S., Vestergaard, M., Storebø, O. J., Sharp, C., & Simonsen, E. (2021). Predictors of dropout among adolescents with borderline personality disorder attending mentalization-based group treatment. *Psychotherapy Research*, *31*(7), 950–961. https://doi.org/10.1080/10503307.2020.1871525

Kasper, L. A., Hauschild, S., Schrauf, L. M., & Taubner, S. (2024). Enhancing mentalization by specific interventions within mentalization-based treatment of adolescents with conduct disorder [Original Research]. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1223040

Kazdin, A. E. (2011). *Single-case research designs: Methods for clinical and applied settings*. Oxford University Press.

Kirtley, O. J., Hiekkaranta, A. P., Kunkels, Y. K., Verhoeven, D., Van Nierop, M., & Myin-Germeys, I. (2019). *The experience sampling method (ESM) item repository*https://doi.org/10.17605/OSF.IO/KG376

Klein Schaarsberg, R. E., Popma, A., Lindauer, R. J. L., & Van Dam, L. (2022). The effects of a virtual Reality–Based Training program for adolescents with disruptive behavior problems on cognitive distortions and treatment motivation: Protocol for a multiple baseline Single-Case experimental design [Protocol]. *JMIR Res Protoc*, *11*(5), e33555. https://doi.org/10.2196/33555

Klein Schaarsberg, R. E., Ribberink, A. Z., Osinga, B., van Dam, L., Lindauer, R. J. L., & Popma, A. (2023). Treatment responsivity in adolescents with disruptive behavior problems: Co-Creation of a virtual Reality–Based Add-On intervention. *JMIR Form Res*, *7*, e46592. https://doi.org/10.2196/46592

Klein Schaarsberg, R. E., van Dam, L., Widdershoven, G. A. M., Lindauer, R. J. L., & Popma, A. (2024). Ethnic representation within virtual reality: A co-design study in A forensic youth care setting. *BMC Digital Health*, *2*(1), 25. https://doi.org/10.1186/s44247-024-00081-0

Klein Schaarsberg, R. E., van Agten, J., de Jong, D., van Haeringen, S. L., van Dam, L., Staaks, J. P. C., Stams, G. J. J. M., Lindauer, R. J. L., & Popma, A. (under review). Mentalization and externalizing behavior from early adolescence to young adulthood: A meta-analysis. *Children and Youth Services Review*. https://doi.org/10.2139/ssrn.4855628

Krasny-Pacini, A., & Evans, J. (2018). Single-case experimental designs to assess intervention effectiveness in rehabilitation: A practical guide. *Annals of Physical and Rehabilitation Medicine*, *61*, 164–179. https://doi.org/10.1016/j.rehab.2017.12.002

Kratochwill, T. R., Hitchcock, J., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2010). Single-case designs technical Documentation. http://ies.ed.gov/ncee/wwc/pdf/wwc_scd.pdf

Kratochwill, T. R., Hitchcock, J. H., Horner, R. H., Levin, J. R., Odom, S. L., Rindskopf, D. M., & Shadish, W. R. (2013). Single-Case intervention research design standards. *Remedial and Special Education*, *34*, 26–38. https://doi.org/10.1177/0741932512452794

Kühn, S., Kugler, D. T., Schmalen, K., Weichenberger, M., Witt, C., & Gallinat, J. (2019). Does playing violent video games cause aggression? A longitudinal intervention study. *Molecular Psychiatry*, *24*(8), 1220–1234. https://doi.org/10.1038/s41380-018-0031-7

Lane, J. D., & Gast, D. L. (2014). Visual analysis in single case experimental design studies: Brief review and guidelines. *Neuropsychological Rehabilitation*, *24*(3–4), 445–463. https://doi.org/10.1080/09602011.2013.815636

Lanovaz, M. J., & Primiani, R. (2023). Waiting for baseline stability in single-case designs: Is it worth the time and effort? *Behavior Research Methods*, *55*(2), 843–854. https://doi.org/10.3758/s13428-022-01858-9

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 & Adolescent Psychology*, *48*(5), 799–810. https://doi.org/10.1080/15374416.2019.1567349

Medland, H., De France, K., Hollenstein, T., Mussoff, D., & Koval, P. (2020). Regulating emotion systems in everyday life: Reliability and validity of the RESS-EMA scale. *European Journal of Psychological Assessment*. https://doi.org/10.1027/1015-5759/a000595

Mestdagh, M., Verdonck, S., Piot, M., Niemeijer, K., Kilani, G., Tuerlinckx, F., Kuppens, P., & Dejonckheere, E. (2023). m-Path: An easy-to-use And highly tailorable platform for ecological momentary assessment And intervention in behavioral research And clinical practice [Technology And Code]. *Frontiers in Digital Health*, *5*. https://doi.org/10.3389/fdgth.2023.1182175

Miller, W. R., & Rollnick, S. (2002). *Motivational interviewing: Preparing people for change* (2nd ed.). Guilford Press.

Miller, S., Duncan, B., & Brown, J. (2003). The outcome rating scale: A preliminary study of the reliability, validity, and feasibility of a brief visual analog measure. *Journal of Brief Therapy*, *2*, 91–100.

Möller, C., Falkenström, F., Larsson, H., M., & Holmqvist, R. (2014). Mentalizing in young offenders. *Psychoanalytic Psychology*, *31*, 84–99. https://doi.org/10.1037/a0035555

Moonen, X., & Kaal, H. (2017). Jeugdigen En Jongvolwassenen Met licht Verstandelijke Beperkingen En criminaliteit. *Justitiële Verkenningen*, *43*, 9–24. https://doi.org/10.5553/JV/016758502017043006002

Myin-Germeys, I., & Kuppens, P. (2022). *The open handbook of experience sampling methodology: A step-by-step guide to designing, conducting, and analyzing ESM studies* (2nd ed.). Center for Research on Experience Sampling and Ambulatory Methods Leuven.

Nas, C. N., Brugman, D., & Koops, W. (2008). Measuring self-serving cognitive distortions with the how I think questionnaire. *European Journal of Psychological Assessment*, *24*(3), 181–189. https://doi.org/10.1027/1015-5759.24.3.181

Nijman, S. A., Veling, W., Greaves-Lord, K., Vos, M., Zandee, C. E. R., Aan het Rot, M., Geraets, C. N. W., & Pijnenborg, G. H. M. (2020). Dynamic interactive social cognition training in virtual reality (DiSCoVR) for people with a psychotic disorder: Single-Group feasibility and acceptability study. *JMIR Ment Health*, *7*(8), e17808. https://doi.org/10.2196/17808

Njardvik, U., Smaradottir, H., & Öst, L. G. (2022). The effects of emotion regulation treatment on disruptive behavior problems in children: A randomized controlled trial. *Research on Child and Adolescent Psychopathology*, *50*(7), 895–905. https://doi.org/10.1007/s10802-022-00903-7

Normand, M. P. (2016). Less is more: Psychologists can learn more by studying fewer people [Opinion]. *Frontiers in Psychology*, *7*, 934. https://doi.org/10.3389/fpsyg.2016.00934

O'Keeffe, S., Martin, P., Goodyer, I. M., Wilkinson, P., Consortium, I., & Midgley, N. (2018). Predicting dropout in adolescents receiving therapy for depression. *Psychotherapy Research*, *28*(5), 708–721. https://doi.org/10.1080/10503307.2017.1393576

Olver, M. E., Stockdale, K. C., & Wormith, J. S. (2011). A meta-analysis of predictors of offender treatment attrition and its relationship to recidivism. *Journal of Consulting and Clinical Psychology*, *79*, 6–21. https://doi.org/10.1037/a0022200

Parkline Interactive, L. L. C. (2019). *Wander [virtual reality application software]*. In (Version 0.3522) https://www.meta.com/nl-nl/experiences/2078376005587859/?utm_source=www.parklineinteractive.com_utm_medium=oculusredirect

Rhee, H., Belyea, M., & Mammen, J. (2017). Visual analogue scale (VAS) as A monitoring tool for daily changes in asthma symptoms in adolescents: A prospective study. *Allergy Asthma Clin Immunol*, *13*, 24. https://doi.org/10.1186/s13223-017-0196-7

Richter, A., Sjunnestrand, M., Strandh, R., M., & Hasson, H. (2022). Implementing school-based mental health services: A scoping review of the literature summarizing the factors that affect implementation. *International Journal of Environmental Research and Public Health*, *19*(6). https://doi.org/10.3390/ijerph19063489

Riise, E. N., Wergeland, G. J. H., Njardvik, U., & Öst, L. G. (2021). Cognitive behavior therapy for externalizing disorders in children and adolescents in routine clinical care: A systematic review and meta-analysis. *Clinical Psychology Review*, *83*, 101954. https://doi.org/10.1016/j.cpr.2020.101954

Rossner, M., & Taylor, H. (2024). The transformative potential of restorative justice: what the mainstream can learn from the margins. *Annual Review of Criminology*, *7*(Volume 7, 2024), 357–381. https://doi.org/10.1146/annurev-criminol-030421-040921

Sauter, F. M., Heyne, D., Blote, A. W., van Widenfelt, B. M., & Westenberg, P. M. (2010). Assessing therapy-relevant cognitive capacities in young people: Development and psychometric evaluation of the self-reflection and insight scale for youth. *Behavioural and Cognitive Psychotherapy*, *38*(3), 303–317. https://doi.org/10.1017/S1352465810000020

Shao, R., & Wang, Y. (2019). The relation of violent video games to adolescent aggression: An examination of moderated mediation effect [Original Research]. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.00384

Shrout, P. E., Stadler, G., Lane, S. P., McClure, M. J., Jackson, G. L., Clavél, F. D., Iida, M., Gleason, M. E. J., Xu, J. H., & Bolger, N. (2018). Initial elevation bias in subjective reports. *Proc Natl Acad Sci U S A*, *115*(1), E15–e23. https://doi.org/10.1073/pnas.1712277115

Smeijers, D., Bulten, E., Buitelaar, J., & Verkes, R. J. (2018). Treatment responsivity of aggressive forensic psychiatric outpatients. *International Journal of Offender Therapy and Comparative Criminology*, *62*(12), 3834–3852. https://doi.org/10.1177/0306624x17747052

Stige, S. H., Barca, T., Lavik, K. O., & Moltu, C. (2021). Barriers and facilitators in adolescent psychotherapy initiated by Adults—Experiences that differentiate adolescents' trajectories through mental health care. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.633663. Original Research.

Suzuki, M. (2023). Victim recovery in restorative justice: A theoretical framework. *Criminal Justice and Behavior*, *50*(12), 1893–1908. https://doi.org/10.1177/00938548231206828

Tate, R. L., Perdices, M., Rosenkoetter, U., McDonald, S., Togher, L., Shadish, W., Horner, R., Kratochwill, T., Barlow, D. H., & Kazdin, A. (2016). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016: Explanation and elaboration. *Archives of Scientific Psychology*, *4*(1), 10.

Tate, R. L., Perdices, M., Rosenkoetter, U., Shadish, W., Vohra, S., Barlow, D. H., Horner, R., Kazdin, A., Kratochwill, T., & McDonald, S. (2016b). The single-case reporting guideline in behavioural interventions (SCRIBE) 2016 statement. *Physical Therapy*, *96*(7), e1–e10.

te Brinke, L. W., Schuiringa, H. D., & Matthys, W. (2021). Emotion regulation and angry mood among adolescents with externalizing problems and intellectual disabilities. *Research in Developmental Disabilities*, *109*, 103833. https://doi.org/10.1016/j.ridd.2020.103833

te Velde, R., Steur, J., & Vankan, A. (2015). *Gaming En gamification voor justitiële inrichtingen*. Wetenschappelijk Onderzoek- en Documentatiecentrum.

Thurmond, V. A. (2001). The point of triangulation. *Journal of Nursing Scholarship*, *33*(3), 253–258. https://doi.org/10.1111/j.1547-5069.2001.00253.x

Van Bockstaele, B., van der Molen, M. J., van Nieuwenhuijzen, M., & Salemink, E. (2020). Modification of hostile attribution bias reduces self-reported reactive aggressive behavior in adolescents. *Journal of Experimental Child Psychology*, *194*, 104811. https://doi.org/10.1016/j.jecp.2020.104811

Van der Helm, G. H., Wissink, I. B., De Jongh, T., & Stams, G. J. (2013). Measuring treatment motivation in secure juvenile facilities. *International Journal of offender Therapy and Comparative Criminology*, *57*(8), 996–1008. https://doi.org/10.1177/0306624X12443798

van Gelder, J. L., De Vries, R. E., Demetriou, A., van Sintemaartensdijk, I., & Donker, T. (2019). The virtual reality scenario method: Moving from imagination to immersion in criminal decision-making research. *Journal of Research in Crime and Delinquency*, *56*, 451–480. https://doi.org/10.1177/0022427818819696

van Linschoten, R. C. A., Knijnenburg, S. L., West, R. L., & van Noord, D. (2023). CastorEDC API: A python package for managing real world data in castor electronic data capture. *Journal of Open Research Software*, *11*(1), 12. https://doi.org/10.5334/jors.436

van Loon, A., Bailenson, J., Zaki, J., Bostick, J., & Willer, R. (2018). Virtual reality perspective-taking increases cognitive empathy for specific others. *PLOS ONE*, *13*(8), e0202442. https://doi.org/10.1371/journal.pone.0202442

Verhoef, R. E. J., Alsem, S. C., Verhulp, E. E., & De Castro, O., B (2019). Hostile intent attribution and aggressive behavior in children revisited: A Meta-Analysis. *Child Development*, *90*(5), e525–e547. https://doi.org/10.1111/cdev.13255

Vermaes, I. P. R., Konijn, C., Jambroes, T., & Nijhof, K. S. (2014). Statische En dynamische Kenmekren Van jeugdigen in JeugdzorgPlus: Een systematische review. *Orthopedagogiek: Onderzoek En Praktijk*, *53*, 278–292.

Voorendonk, E. M., De Jongh, A., Rozendaal, L., & Van Minnen, A. (2020). Trauma-focused treatment outcome for complex PTSD patients: Results of an intensive treatment programme. *European Journal of Psychotraumatology*, *11*(1), 1783955. https://doi.org/10.1080/20008198.2020.1783955

Walker, S. G., & Carr, J. E. (2021). Generality of findings from single-case designs: It's not all about the N. *Behav Anal Pract*, *14*(4), 991–995. https://doi.org/10.1007/s40617-020-00547-3

Ward, T., Day, A., Howells, K., & Birgden, A. (2004). The multifactor offender readiness model. *Aggression and Violent Behavior*, *9*, 645–673.

Wei, M., Liu, Y., & Chen, S. (2022). Violent video game exposure and problem behaviors among children and adolescents: The mediating role of deviant peer affiliation for gender and grade differences. *International Journal of Environmental Research and Public Health*, *19*(22). https://doi.org/10.3390/ijerph192215400

Weisz, J. R., Chorpita, B. F., Frye, A., Yi Ng, M., Lau, N., Bearman, K., Ugueto, S., Langer, A. M., D. A., & Hoagwood, K. E. (2011). Youth top problems: Using idiographic, consumer-guided assessment to identify treatment needs and to track change during psychotherapy. *Journal of Consulting and Clinical Psychology*, *79*, 369–380. https://doi.org/10.1037/a0023307

Weisz, J. R., Kuppens, S., Ng, M. Y., Vaughn-Coaxum, R. A., Ugueto, A. M., Eckshtain, D., & Corteselli, K. A. (2018). 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*(2), 216–237. https://doi.org/10.1177/1745691618805436

World Medical Association. (2013). World medical association declaration of Helsinki: Ethical principles for medical research involving human subjects. *Journal of the American Medical Association*, *310*(20), 2191–2194. https://doi.org/10.1001/jama.2013.281053

Yirmiya, N., Erel, O., Shaked, M., & Solomonica-Levi, D. (1998). Meta-analyses comparing theory of mind abilities of individuals with autism, individuals with mental retardation, and normally developing individuals. *Psychological Bulletin*, *124*, 283–307.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.