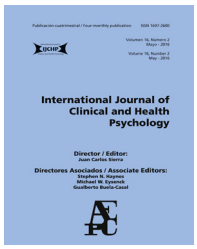
International Journal of Clinical and Health Psychology 21 (2021) 100226

International Journal of Clinical and Health Psychology

www.elsevier.es/ijchp



ORIGINAL ARTICLE

Effects of cognitive bias modification for interpretation on hostile interpretation bias and self-reported aggression in juvenile delinquents

Zhihong Ren[a,1], **Ziyi Zhao**[a,1], **Xianglian Yu**[a,b,*], **Lin Zhang**[a], **Xu Li**[a]

[a] *Key Laboratory of Adolescent Cyberpsychology and Behavior (CCNU), Ministry of Education, Key Laboratory of Human Development and Mental Health of Hubei Province, School of Psychology, Central China Normal University, China*
[b] *Department of Education, Jianghan University, China*

Received 11 September 2020; accepted 4 January 2021

KEYWORDS
Interpretation bias modification;
Aggression;
Juvenile delinquents;
Hostile interpretation bias;
Experiment

Abstract
Background/Objective: The social information processing model holds that aggressive behavior is closely related to the hostile interpretation of ambiguous social cues, suggesting the possibility that an intervention that reduces hostile interpretations could reduce aggression. This study in Mainland China evaluated the remediating effects of cognitive bias modification for interpretation (CBM-I) on the hostile interpretation bias and self-reported aggressive behaviors of male juvenile delinquents, taking into account initial hostile interpretation bias as a possible moderator of the intervention effect.
Method: Fifty-six male juvenile delinquents aged 16-18 were recruited and randomly assigned to the CBM-I group (n = 28) or the Waiting-List group (n = 28). Interpretation bias and self-reported aggressive behavior were assessed at pre-test and post-test.
Results: The positive interpretations of participants in the CBM-I group were significantly increased compared with participants in the Waiting-List group. The intervention effect of CBM-I on self-reported physical aggression was significant only for juvenile delinquents with high pre-test hostile interpretation bias.
Conclusions: CBM-I can significantly improve the positive interpretation bias of juvenile delinquents, and reduce the self-reported physical aggression for some male juvenile delinquents. The results have implications for providing low-cost and high-efficiency intervention for juvenile delinquents' self-reported aggression behavior.

* Corresponding author: School of Psychology, Central China Normal University, NO.152 Luoyu Road, Wuhan, Hubei, 430079 PR China.
E-mail address: psyyu@jhun.edu.cn (X. Yu).
[1] These authors contributed equally to this work.

https://doi.org/10.1016/j.ijchp.2021.100226

PALABRAS CLAVE
Modificación del sesgo de interpretación;
Agresión;
Delincuentes juveniles;
Sesgo de interpretación hostil;
Experimento

Efectos de la modificación del sesgo cognitivo sobre la interpretación del sesgo de interpretación hostil y agresión autoinformada en delincuentes juveniles

Resumen
Antecedentes/Objetivo: El modelo de procesamiento de información social sostiene que el comportamiento agresivo está relacionado con la interpretación hostil de señales sociales ambiguas, lo que sugiere que una intervención que reduzca las interpretaciones hostiles podría reducir la agresión. Se evaluaron los efectos de la modificación del sesgo cognitivo para la interpretación (CBM-I, por sus siglas en inglés) sobre el sesgo de interpretación hostil y conductas agresivas autoinformadas de delincuentes juveniles, teniendo en cuenta el sesgo de interpretación hostil inicial como posible moderador del efecto de la intervención.
Método: Cincuenta y seis delincuentes juveniles varones (16-18 años) se asignaron al azar al grupo CBM-I (n = 28) o al grupo lista de espera (n = 28). El sesgo de interpretación y el comportamiento agresivo autoinformado se evaluaron en pretest y postest.
Resultados: Las interpretaciones positivas de los participantes en el grupo CBM-I aumentaron significativamente en comparación con el grupo Lista de espera. El efecto de la intervención de CBM-I sobre la agresión autoinformada fue significativo solo para delincuentes con un alto sesgo de interpretación hostil en el pretest.
Conclusiones: CBM-I puede mejorar significativamente el sesgo de interpretación positiva de delincuentes juveniles y reducir la agresión física autoinformada en algunos delincuentes juveniles masculinos.

As a prominent problem worldwide, violent crimes committed by juveniles create a large economic burden and general destruction of the fabric of society (Krug et al., 2002). A survey by the Office of Juvenile Justice and Delinquency Prevention of the United States found that juveniles (ages 12-17) were involved in about 16% of serious violent crimes annually between 2007 and 2017, with the rate fluctuating between 9.70% and 22.30%. The number of juvenile homicide offenders in the U.S. who used a firearm increased by 82% from 2013 to 2018. According to the Chinese National Bureau of Statistics, there were 34 thousand youth classified as juvenile delinquents in 2018, and juvenile delinquents (ages 14-17) accounted for 2.41% to 6.78% of the total number of criminals between 2010 and 2018. Violent crime, including aggravated assault, rape, and robbery, is the main form of juvenile crime in China and accounts for 73.38% of all types of juvenile crime in 2017 (Lu et al., 2018). Aggressive behavior represents an important risk factor for criminality in adulthood (Assink et al., 2015). Existing research on aggressive behavior has focused mostly on adult criminals or general adolescents, and empirical evidence on interventions targeting aggression in high-risk youths, namely juvenile delinquents, is still lacking. In China, juvenile delinquents refer to criminals aged between 14 and 18 years-old who are considered old enough to be responsible for their criminal behaviors. Because of their young age, they won't be sent to prison but juvenile correctional institutions. That is, instead of incarceration, they receive educational rehabilitation in juvenile correctional institutions. Considering the severity of the status quo of juvenile violent crime and the scarcity of previous research, it is necessary to develop interventions to reduce aggression in this population.

The generation and development of aggressive behavior both are associated with biased social cognitive processes, especially the hostile interpretation of ambiguous social cues (Lochman & Dodge, 1994). The Social Information Processing (SIP) model maintains that the encoding and interpretation of social cues are the basis of subsequent emotional and behavioral responses. The hostile interpretation bias can lead to an unwarranted sense of being threatened, leading to negative emotions and finally aggressive behavior (Crick & Dodge, 1994). Similarly, the Integrative Cognitive Model maintains that hostile interpretation bias can increase aggression by inducing anger (Wilkowski & Robinson, 2008). Consistent with these models, two recent meta-analyses showed that the hostility bias when interpreting facial expressions and ambiguous situations is significantly associated with aggression in samples of both children (Verhoef et al., 2019) and adults (Tuente et al., 2019).

Taking into account the positive association between hostile interpretation bias and aggressive behavior, correcting this bias has the potential to reduce aggression. Cognitive bias modification for interpretation (CBM-I) is such an intervention aimed at the interpretation process (Koster et al., 2009). Unlike traditional therapies that rely on conscious introspection and reflection, CBM-I aims to directly change interpretations with a down-top approach. The therapy guides the individual to use a positive interpretation style repeatedly and to make benign attributions in ambiguous situations, to reduce hostile interpretation bias in turn (Jones & Sharpe, 2017). Furthermore, CBM-I can also be well combined with a computer, providing a low-cost and high-efficiency intervention. Studies have shown

that CBM-I was effective in reducing body dissatisfaction in a community sample (Dietel et al., 2020), remedying appearance and self-worth biases in eating disorder (Matheson et al., 2019), and reducing negative interpretation bias in people with state social anxiety (Yeung & Sharpe, 2019), but no research has been conducted to intervene the juvenile delinquents' aggressive behaviors with CBM-I. Given the close relationship between hostile interpretation bias and aggressive behavior, we believe that the CBM-I can reduce the aggression level by remediating the hostile interpretation bias of juvenile delinquents.

Previous studies have explored the training effect of CBM-I on negative interpretation bias, but the results were not consistent. One study showed that CBM-I could effectively reduce the negative interpretation bias in individuals with social anxiety (Yeung & Sharpe, 2019); the other reported that CBM-I has no significant effect on the interpretation bias of individuals with anxiety disorders (MacDonald et al., 2020). In this regard, it has been pointed out that the consistency/inconsistency of the materials and tasks used in the intervention and outcome evaluation may have a confounding effect (Hertel & Mathews, 2011), the more similar the two sets of materials and tasks, the larger the estimated effect size (Gonsalves et al., 2019). This study extended previous studies by evaluating the effectiveness of CBM-I intervention across paradigms and materials, including several different tasks to measuring the effectiveness of CBM-I in reducing hostility bias in the interpretation of ambiguous situations, facial expressions, and intentions.

What's more, the participants' baseline hostile interpretation bias might affect the training effects of CBM-I. CBM-I requires two premises for its effectiveness: the maladaptive interpretation bias exhibited by individuals, and the close relationship between maladaptive interpretation pattern and the follow up maladaptive emotion and behavior (Jones & Sharpe, 2017). According to the premises, comparing with the individuals with low levels of initial hostile interpretation bias, those with high levels of initial hostile interpretation bias may be more consistent with the theoretical premises of CBM-I, and obtain greater benefits from the intervention. Consistent with this view, researchers found that the effect of CBM-I was enhanced for participants with a severe baseline level of interpretation bias (Beard et al., 2019; Micco et al., 2014; Salemink & Wiers, 2011).

The present study tested the effect of CBM-I on hostile interpretation bias and self-reported aggressive behavior in a sample of youth who had been labeled juvenile delinquents. The participants' baseline hostile interpretation bias was also taken into account as a moderator of the treatment effect. We hypothesized that at the end of the study, participants in the CBM-I group would show less hostile interpretation bias, higher positive interpretation bias, and lower self-reported aggression than those in the Waiting-List group. In addition, a larger intervention effect would be found among individuals with high initial hostile interpretation scores.

Method

Participants

Fifty-six male adolescents (aged 16-18 years; *M* = 17.52, *SD* = 0.57) who had been identified as juvenile delinquents were recruited from a juvenile correctional institution in an urban area of South China. Exclusion criteria were juvenile delinquents with (a) history of mental illness or attention disorders; (b) current treatment with psychotherapy or psychotropic medication; (c) difficulty in reading or color recognition; (d) intellectual disabilities. The demographic and crime information of participants was presented in Table 1.

Instruments

The Chinese version of the Buss and Perry Aggression Questionnaire (AQCV) was used to measure the participants' aggressiveness (Li et al., 2011). This scale consists of thirty items constituting five subscales: Physical Aggression (AQCV-PA; e.g., ''Once in a while, I can't control the urge to strike another person''), Verbal Aggression (AQCV-VA; e.g., ''I can't help getting into arguments when people disagree with me''), Anger (AQCV-A; e.g., ''I flare up quickly but get over it quickly''), Hostility (AQCV-H; e.g., ''I am suspicious of overly friendly strangers''), Self-Directed Aggression (AQCV-SA; e.g., ''When I'm upset, I think about hurting myself''). Participants were instructed to rate each item from 1 (*completely untrue)* to 5 (*completely true)*. AQCV has satisfactory reliability in Chinese adolescents (Liu et al., 2009). In the present study, the Cronbach's α of the questionnaire was .94 in the pre-test session, and .93 in the post-test session.

The Chinese version of the Ambiguous Intentions Hostility Questionnaire (AIHQ) was used to measure interpretation biases when others' intentions are ambiguous (Chen et al., 2012). The questionnaire includes fifteen short vignettes that describe negative interpersonal events in which the motives of one character are not clear. Participants were asked to read each vignette and imagine the scenario happening to themselves (e.g., ''You walk past a bunch of teenagers at a mall and you hear them start to laugh''). After each vignette, the participants were asked to respond to an open-ended question, three Likert-scale items, and then another open-ended question. For the first open-ended question, participants were asked to write down the reason why the other person (or persons) behaved the way they did, which was later coded by raters to compute the Hostile Index (AIHQ-H). For the Likert scale items, participants were asked to rate whether the other person (or persons) performed the action on purpose (1 = *definitely no* to 6 = *definitely yes*), how angry it would make them feel (1 = *not at all angry* to 5 = *very angry*), and how much they would blame the other person (or persons) (1 = *not at all* to 5 = *very much*). The ratings on the Likert scale items were summed to create the Blame Index (AIHQ-B). Finally,

Table 1 Demographic and crime information of participants.

Variables	CBM-I group (n = 27)	Waiting-List group (n = 25)
Age (years)	17.52 (0.75)	17.56 (0.58)
Education (years)	8.89 (1.55)	8.64 (1.32)
Type of crime		
Property crime	17 (63.00%)	14 (56%)
Violent crime	9 (33.30%)	11 (44%)
Other	1 (3.70%)	0 (0%)
Assigned time in correctional institution		
Within one year	2 (7.40%)	1 (4%)
One to three years (including three years)	21 (77.80%)	19 (76%)
Three to ten years (including ten years)	4 (14.80%)	5 (20%)

the participants were asked to respond to the second open-ended question by writing down how he/she would respond to the situation, which was later coded by raters to compute an Aggression Index (AIHQ-A). Four graduate students majoring in psychology independently rated the responses to the two open-ended questions on the AIHQ. The raters were trained by providing examples of responses that would receive high and low scores on hostility and aggression. The raters used a 5-point Likert scale to evaluate the hostile level for open-response question 1 (1 = *not at all hostile* to 5 = *very hostile*), and the aggressive level for open-response question 2 (1 = *not at all aggressive* to 5 = *very aggressive*). The ratings were averaged to create the Hostility Index and the Aggression Index (Combs et al., 2007). In this study, the Cronbach's α for the self-reported Blame Index (AIHQ-B) was .89 at pre-test and .97 at post-test session. The average intraclass correlation coefficients (ICCs) were high for both the Hostility Index (AIHQ-H; .81 at pre-test and .83 at post-test) and the Aggression Index (AIHQ-A; .84 at pre-test and .85 at post-test).

The Word Sentence Association Paradigm-Hostility (WSAP) task was utilized to measure the participants' hostile interpretation bias in evaluating ambiguous situations (Dillon et al., 2016). E-Prime 2.0 (Psychology Software Tools, Sharpsburg, PA, USA) was used to program and administer the task. Ten practice trials preceded the formal experiment, which consisted of sixty-six trials. First, a gaze point was displayed in the center of the screen, then an ambiguous sentence (such as S̈omeone is in your waÿ), and an interpretation for the sentence, positive (Ḧe/She is unawarë) or hostile (Ḧe/She is inconsideratë), were presented. The participants were asked to indicate the similarity of the interpretation with their own interpretation of the sentence (1 = *not similar at all* to 6 = *completely similar*). This task had 16 ambiguous sentences and 32 interpretation phrases (16 positive, 16 hostile) in total. Some changes were made to the task items so they better reflected the juvenile delinquents' daily life. For example, T̈he boss asks you to do some extra workẅas replaced with T̈he institution guard asks you to do some extra work.T̈he average similarity rating for all positive interpretations was used as the Positive Interpretation score (WSAP-P), and the average similarity rating for all hostile interpretations was used as the Hostile Interpretation score (WSAP-H).

The Hostile Interpretation Bias Task (HIBT) was used to measure hostile interpretation bias towards facial expressions (Smeijers et al., 2017). FaceGen3.4 (http://FaceGen.com; Singular Inversions, 2009) provided photographs of four males and four females expressing six emotional facial expressions (anger, fear, disgust, happiness, surprise, sadness), and a neutral facial expression used as a benchmark. In addition, each emotional expression was deformed five times using WinMorph 3.01 to alter its emotional intensity: of 20%, 40%, 60%, 80%, and 100% intense. In all there were 240 facial expression pictures. The experimental program was compiled using E-Prime 2.0 (Psychology Software Tools, Pittsburgh), including the practice stage (16 trials only with happy faces and angry faces with 100% intensity) and formal experiment (240 trials). A gaze point was firstly displayed in the center of the screen, then a face picture was presented, and the participant was asked to judge whether the facial expression was hostile or not, and to register their judgment by pushing a corresponding button as soon as possible. The percentage of the participant's ''hostile'' reactions towards facial expressions was calculated as the hostile interpretation bias score.

Procedure

A randomized controlled trial was set up to investigate the effect of CBM-I. The participants firstly completed the pre-test questionnaire assessments and pre-test behavioral tasks on the computer. The pre-test assessments took about thirty minutes to complete. The participants were then randomly assigned to the CBM-I group or the Waiting-List group, with 28 persons in each group. The CBM-I group completed four intervention tasks that took place once a week with an interval of seven days over the course of one month; the Waiting-List group did not receive any intervention. After the intervention, all participants in both groups completed the same pre-test assessments as post-test outcome measures. As all participants were recruited from the same juvenile correctional institution. Given the possibility that the participants would share information about the task, the purpose of the study was disguised. Participants were told that the experiment was aimed at recording changes in men-

tal status, and that test intervals could be one week or one month respectively for the two groups, and the tests might differ to some extent for participants in different groups.

This project was approved by the Ethics Committee for the Humanities and Social Sciences at Fuzhou University (EC2018021) and has been registered (osf.io/gyq35). All participants participated in the study voluntarily and were told that they could withdraw from the study at any time for any reason. In addition, we promised the participants that the relevant research data and results would not be disclosed to the juvenile correctional institution. Informed consent was obtained from all participants and their guard unit (juvenile correctional institution).

Intervention

An adapted version of the computerized CBM-I was employed to modify the hostile interpretation bias of juvenile delinquents (Brettschneider et al., 2015). The intervention materials were 36 ambiguous scenarios, three options for how to interpret the scenario, and corresponding feedback about which option the participant chose. The screen first presented an ambiguous scenario and three interpretations of the scenario (a positive, a neutral, and a hostile interpretation), and the participant was asked to choose the option most similar to his/her interpretation, and then the feedback was given based on the participant's option. The feedback corresponding to the positive or neutral interpretation included words of encouragement and reinforcement, while the feedback corresponding to the hostile option included information about another way to interpret the scenario from a different perspective. The scene materials and feedback used in the intervention were compiled by four graduate students majoring in psychology and reviewed by the management staff of the juvenile correctional institution to ensure that they were in line with the life scenes of the juvenile delinquents.

The treatment was programmed using E-prime 2.0 (Psychology Software Tools, Pittsburgh). There were four treatment sessions. The first two sessions (ten scenes each) provided feedback to the participants' choice of how to respond to the scenario in the vignette. The feedback was presented on the screen for 20 seconds. The other two treatment sessions (eight scenes each) presented guidelines for 20 seconds to ask the participants to think of the reason for their choice and its consequence, and to think of the ambiguous scenarios from multiple perspectives.

Statistical analyses

For the WSAP, two participants did not complete the pre-test, and another two did not finish the post-test. For the HIBT, three participants did not complete the pre-test task, and one evaluated all stimuli as hostile in the pre-test task and did not complete the post-test. Three participants with missing data in the pre-test and one participant with extreme reaction were excluded. Finally, 27 participants in the CBM-I group and 25 participants in the Waiting-List group were included in the final data analysis. Figure 1 presents the CONSORT (Consolidated Standards of Reporting Trials) diagram of the current research.

Intention to treat analysis (ITT) was applied to keep all the participants in the randomly assigned groups. For participants with missing post-test data, their pre-test data was used to fill in. Mixed Linear Modeling (MLM) was employed to analyze the influence of the intervention on the main variables (hostile interpretation bias scores, positive interpretation bias scores, and self-reported aggression). MLM is suitable for repeated measures and fits randomized controlled trial research well. Age and education were controlled, then time (pre-test, post-test), group (CBM-I group, Waiting-List group), and their interaction effect were set as fixed effects, the participant was set as a random effect. Interpretation bias scores (WSAP, AIHQ, HIBT) and self-reported aggression (AQCV and its five subscales) were set as dependent variables. The restricted maximum likelihood method was used to fit the model data, and the autoregressive covariance structure (AR1) was chosen for model analysis. In addition, the hierarchical regression analysis was utilized to explore the moderating effect of pre-test interpretation bias scores on the intervention effect. Both the MLM and the moderating effect analysis were conducted with SPSS 20.0.

Results

Group differences at pre-test

Independent sample *t*-tests were used to compare the two groups at pre-test on the demographic variables and main outcome variables (Table 2). At pre-test the CBM-I group had significantly higher scores on the AIHQ-H than the Waiting-List group, t (50) = -2.44, p = .018, Cohen'd = 0.68, 95% CI [-8.64, -0.85]. At pre-test the Waiting-List group made hostile judgments of the high intensity fear expressions (80% fear intensity, pre-HIBT-80-F; 100% fear intensity, pre-HIBT-100-F) that were significantly higher than those of the CBM-I group (pre-HIBT-80-F: t (50) = 2.29, p = .027, Cohen'd = -0.64, 95% CI [0.02, 0.34]; pre-HIBT-100-F: t (50) = 2.14, p = .036, Cohen'd = -0.60, 95% CI [0.01, 0.34]). At pre-test there were no significant group differences on any of the other measures.

Treatment effects on hostile interpretation bias

An MLM was constructed to test the effect of the CBM-I intervention on the interpretation bias of juvenile delinquents, the results of which are shown in Table 3. In between-group analyses, MLM on the WSAP-P revealed a significant effect of the Time × Group interaction. Follow-up analyses showed that there was no significant difference between the two groups before the intervention, t (50) = 1.18, p = .244; after the intervention, the CBM-I group's WSAP-P score was significantly higher than that of the Waiting-List group, t (50) = -2.24, p = .030, Cohen'd = 0.62, 95% CI [-1.41, -0.08]. In within-group analyses, there was significant improvement from pre- to post-test WSAP-P scores in the CBM-I group, t (50) = -2.85, p = .008, Cohen'd = 0.50, 95% CI [-0.91, -0.15], but not for the Waiting-List group, t (50) = 2.06, p = .050, Cohen'd = -0.48, 95% CI [-0.001, 1.05].

MLM on the WSAP-H also revealed a significant effect of the Time × Group interaction. For the WSAP-H scores,

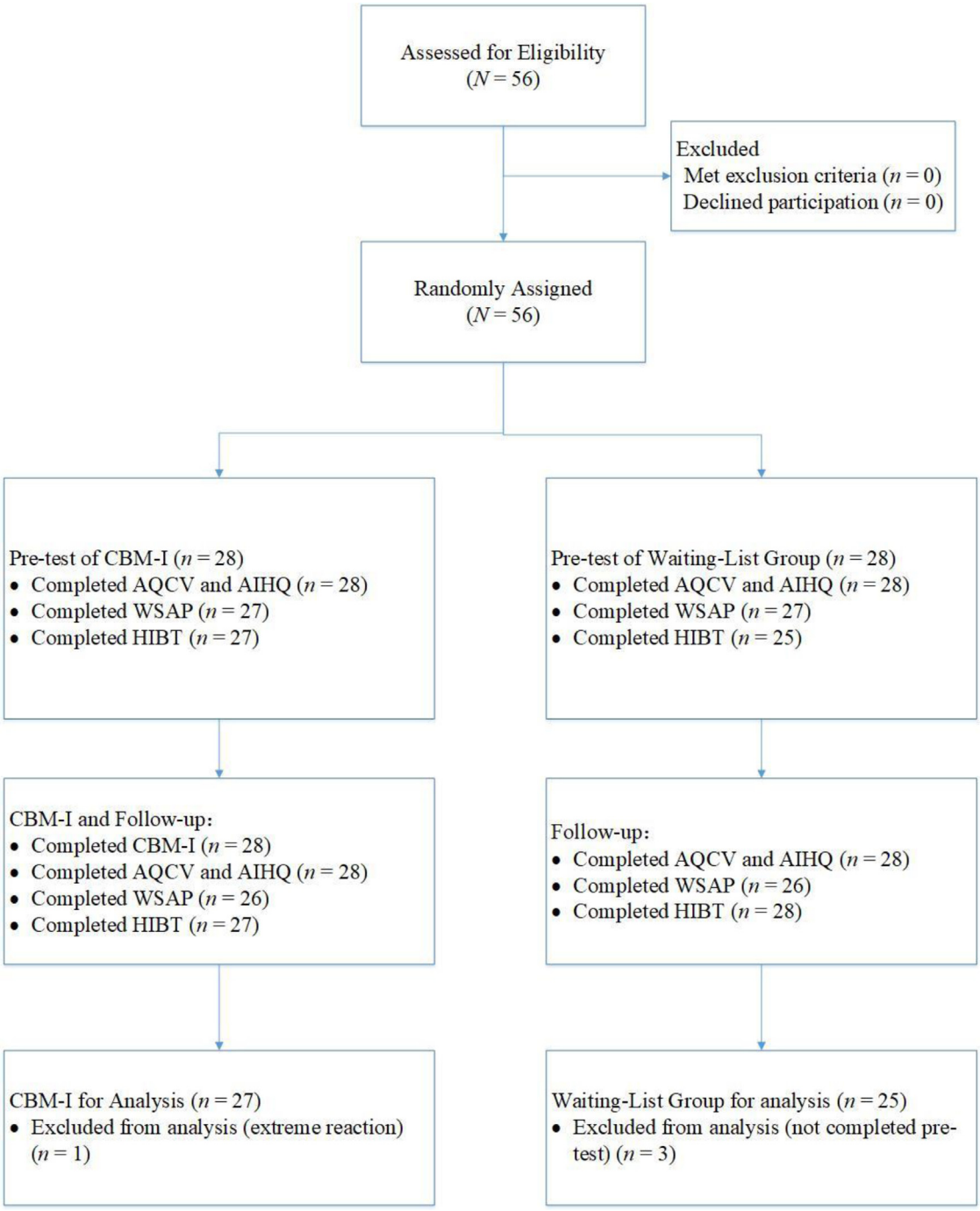


Figure 1 CONSORT diagram illustrating the research process.
Note. CONSORT: Consolidated Standards of Reporting Trials. CBM-I: CBM-I group; AQCV: Buss-Perry Aggression Questionnaire; AIHQ: Ambiguous Intentions Hostility Questionnaire; WSAP: Word Sentence Association Paradigm; HIBT: The Hostile Interpretation Bias Task.

the difference between the two groups before the intervention was not significant, t (50) = -0.27, p = .789; after the intervention, the WSAP-H of the CBM-I group was significantly higher than that of the Waiting-List group, t (50) = -2.32, p = .024, Cohen'd = 0.64, 95% CI [-1.58, -0.11]; the analysis of the difference between the pre- and post-test within the group indicated that, for the CBM-I group, the difference between the pre- and post-test WSAP-H was not significant, t (50) = -1.18, p = .248, Cohen'd = -0.16, 95% CI [-0.57, 0.15]; for the Waiting-List group, the post-test WSAP-H score was significantly lower than the pre-test score, t (50) = 2.41, p = .024, Cohen'd = -0.45, 95% CI [0.07, 1.01]. The treatment effects on AIHQ and HIBT were not significant.

Table 2 Descriptive statistics for pre-test measures.

Variables	Waiting-List group (*n* = 25)		CBM-I group (*n* = 27)		Difference	
	M	*SD*	*M*	*SD*	*t*	*p*
WSAP						
WSAP-P	4.23	0.76	3.91	1.11	1.19	.24
WSAP-H	2.95	1.26	3.05	1.25	−0.27	.79
AIHQ						
AIHQ-B	101.68	25.52	113.04	24.24	−1.65	.11
AIHQ-A	33.84	8.50	33.89	7.59	−0.02	.98
AIHQ-H	28.56	7.22	33.48	7.27	−2.45	.02
HIBT						
Angry (20% - 100%)	0.14-0.84	0.17-0.27	0.10-0.89	0.12-0.24	0.38-1.84	.07-.71
Disgusted (20% - 100%)	0.15-0.83	0.18-0.25	0.09-0.82	0.12-0.29	0.04-1.30	.20-.96
Happy (20% - 100%)	0.10-0.20	0.13-0.24	0.09-0.14	0.16-0.20	0.10-1.45	.15-.92
Surprised (20% - 100%)	0.11-0.20	0.15-0.22	0.06-0.13	0.12-0.20	1.30-1.73	.09-.20
Sad (20% - 100%)	0.12-0.16	0.15-0.17	0.08-0.14	0.12-0.19	0.13-1.58	.30-.91
Fear (20% - 60%)	0.14-0.28	0.23-0.30	0.07-0.15	0.12-0.22	1.40-1.77	.08-.17
Fear 80%	0.39	0.33	0.21	0.23	2.29	.03
Fear 100%	0.45	0.32	0.27	0.28	2.14	.04
AQCV						
AQCV	73.80	22.18	76.33	24.37	−0.39	.70
AQCV-PA	20.00	7.08	21.07	7.69	−0.52	.60
AQCV-VA	12.28	3.86	12.56	4.11	−0.25	.80
AQCV-A	14.68	5.82	15.56	6.30	−0.52	.61
AQCV-H	15.84	5.98	15.48	5.87	0.22	.83
AQCV-SA	11.00	4.18	11.67	4.07	−0.58	.56

Note. *t* in HIBT (The Hostile Interpretation Bias Task) is absolute value. WSAP: Word Sentence Association Paradigm-Hostility; WSAP-P: Word Sentence Association Paradigm-Positive Interpretation; WSAP-H: Word Sentence Association Paradigm-Hostile Interpretation; AIHQ: Ambiguous Intentions Hostility Questionnaire; AIHQ-B: blame bias score of AIHQ; AIHQ-A: aggression bias score of AIHQ; AIHQ-H: hostile bias score of AIHQ; HIBT: The Hostile Interpretation Bias Task; AQCV: Buss-Perry Aggression Questionnaire, The AQCV subscale scores are as follows. AQCV-PA: physical aggression; AQCV-VA: verbal aggression; AQCV-A: anger; AQCV-H: hostility; AQCV-SA: self-directed aggression. The same below.

Treatment effects on self-reported aggressive behavior and the moderating effect of initial interpretation bias

The results of MLM indicated that the intervention did not significantly reduce self-reported aggression (Table 3). Considering that the intervention effect on the self-reported aggression level may have been affected by the individual's initial interpretation bias, and the Time × Group interaction term had a significant effect on WSAP-P and WSAP-H, stepwise regression analysis was employed to further investigate the moderating effects of initial scores on the WSAP-P and WSAP-H on the intervention effect for self-reported aggression.

To prevent multicollinearity, the control variables and the moderating variables were first centralized, and the groups were dummy encoded (Waiting-List group = 0, CBM-I group = 1). The changes in the AQCV total score and its five subscale scores were entered as dependent variables; age and education level were entered as covariates; group, pre-test scores on the WSAP-P / WSAP-H, and the interactions between group and each pre-test score, were entered as predictor variables. The multicollinearity test found that the tolerance of each model was greater than 0.431, and each variance expansion factor was less than 2.318, which showed that there was no serious multicollinearity problem.

The correlation matrix showing correlations among all study variables is presented in Table 4. The results showed that the interaction between the Group and WSAP-H significantly predicted changes in AQCV-PA (Table 5). The simple slopes test was used to further investigate the interaction (Figure 2). For participants with a low level of pre-test hostile interpretation on the WSAP-H, the △AQCV-PA in the CBM-I group was not significantly different from that of the Waiting-List group (*simple slope* = 1.62, *t* = 0.96, *p* = .342); for participants with a high level of pre-test hostile interpretation on the WSAP-H, the △AQCV-PA in the CBM-I group was significantly smaller than that of the Waiting-List group (*simple slope* = -3.95, *t* = -2.32, *p* = .024).

Discussion

The present study tested the intervention effects of CBM-I on the hostile interpretation bias and self-reported aggression in a sample of Chinese youth who had been labeled juvenile delinquents. The results indicated that CBM-I can effectively improve the positive interpretation scores (WSAP-P) of juvenile delinquents for ambiguous situations. The intervention effect on self-reported aggression was more complex.

Table 3 Descriptive statistics for post-test measures and mixed linear model.

Variables	Waiting-List group (n = 25)		CBM-I group (n = 27)		Mixed Linear Model		
	M	*SD*	*M*	*SD*	Time	Group	Time × Group
WSAP							
WSAP-P	3.70	1.36	4.30	1.10	0.00	0.70	11.46***
WSAP-H	2.41	1.17	3.25	1.43	1.40	2.06	6.99**
AIHQ							
AIHQ-B	99.44	31.97	116.89	44.77	0.03	3.37	0.44
AIHQ-A	32.64	8.91	34.48	9.57	0.07	0.20	0.62
AIHQ-H	27.52	7.32	32.26	9.29	1.39	6.11*	0.01
HIBT							
Angry (20% - 100%)	0.22-0.80	0.21-0.31	0.17-0.86	0.23-0.31	0.06-2.86	0.09-2.85	0.01-0.47
Disgust (20% - 100%)	0.16-0.79	0.24-0.31	0.16-0.82	0.24-0.39	0.07-0.96	0.03-0.31	0.23-2.22
Happy (20% - 100%)	0.13-0.19	0.17-0.23	0.11-0.17	0.27-0.32	0.07-1.20	0.01-0.92	0.01-1.03
Surprise (20% - 100%)	0.14-0.22	0.24-0.63	0.12-0.17	0.28-0.33	0.05-1.55	0.61-1.56	0.17-1.84
Sad (20% - 100%)	0.71-1.00	0.19-0.27	0.13-0.16	0.28-0.33	0.26-3.05	0.05-0.71	0.01-0.97
Fear (20% - 80%)	0.18-0.31	0.27-0.32	0.13-0.19	0.29-0.36	0.02-2.94	1.16-3.89	0.01-0.82
Fear 100%	0.42	0.36	0.25	0.34	0.59	4.42*	0.71
AQCV							
AQCV	72.68	21.90	77.22	26.12	0.01	0.30	0.45
AQCV-PA	19.64	7.20	19.48	7.31	2.37	0.06	0.95
AQCV-VA	11.76	3.56	13.81	7.41	0.39	0.85	2.24
AQCV-A	15.08	5.41	15.85	5.72	0.46	0.29	0.01
AQCV-H	15.36	5.99	16.04	6.48	0.01	0.01	2.05
AQCV-SA	10.84	4.02	12.04	4.54	0.05	0.76	0.31

Note. ***p < .001, **p < .01, *p < .05.

Table 4 Correlations among all study variables in the full sample at pre-test and post-test.

Variable	1	2	3	4	5	6	7	8	9	10
1 Age	-	-.11	.18	-.16	-.07	-.16	.05	-.18	.04	.07
2 Education	-.11	-	-.16	-.05	.00	-.08	-.04	.05	.06	.02
3 WSAP-P	.02	.00	-	-.01	.00	.00	.11	.00	-.04	-.03
4 WSAP-H	-.21	-.01	.24	-	.43**	.33*	.38**	.34*	.46**	.29*
5 AQCV	-.10	.01	-.06	.24	-	.89**	.83**	.87**	.82**	.80**
6 AQCV-PA	-.17	.01	-.17	.06	.86**	-	.75**	.76**	.57**	.58**
7 AQCV-VA	-.20	.05	.05	.34*	.75**	.54**	-	.73**	.56**	.52**
8 AQCV-A	-.07	.07	-.07	.05	.85**	.75**	.50**	-	.58**	.57**
9 AQCV-H	.09	-.01	-.11	.27	.84**	.56**	.55**	.62**	-	.78**
10 AQCV-SA	-.07	-.11	.10	.34*	.83**	.60**	.49**	.66**	.77**	-

Note. N = 52. Correlation coefficients of the post-test variables are below the diagonal line; correlation coefficients of pre-test variables are above the diagonal line. WSAP-P: Word Sentence Association Paradigm-Positive Interpretation; WSAP-H: Word Sentence Association Paradigm-Hostile Interpretation; AQCV: Buss-Perry Aggression Questionnaire. The AQCV subscale scores are as follows. AQCV-PA: physical aggression; AQCV-VA: verbal aggression; AQCV-A: anger; AQCV-H: hostility; AQCV-SA: self-directed aggression. ***p < .001, **p < .01, *p < .05.

The benefit of treatment for reducing self-reported aggression was moderated by baseline hostile interpretation bias. Specifically, for individuals with high initial hostile interpretation scores, the intervention program effectively reduced their self-reported physical aggression level; for individuals with low initial hostile interpretation scores, the difference in self-reported physical aggression level between the CBM-I group and the Waiting-List group was not significant. This is the first study to evaluate the remediating effects of cognitive bias modification for interpretation (CBM-I) on the hostile interpretation bias and self-reported aggressive behaviors of male juvenile delinquents, taking into account initial hostile interpretation bias as a possible moderator of the intervention effect.

This study found that CBM-I can prompt juvenile delinquents to make more positive interpretations of ambiguous situations, which is consistent with previous CBM-I studies focused on undergraduate students (Hawkins & Cougle, 2013), adults (Almoghrabi et al., 2018), and aggressive children (Vassilopoulos et al., 2015). Furthermore, this study supports the effectiveness of the CBM-I intervention when applied with juvenile delinquents. Providing immediate

Table 5 Stepwise regression analysis of aggression level on group and pre-test WSAP-H.

Predictors		△ AQCV-PA				
	B	β	*t*	95% CI	$\triangle R^2$	
Step1						.00
	Education	−0.50	-.07	−0.55	[-2.33, 1.33]	
Step2						.10
	Group	−1.08	-.12	−0.90	[-3.47, 1.32]	
	T1-WSAP-H	0.11	.03	0.15	[-1.29, 1.50]	
Step 3						**.09***
	Group × T1-WSAP-H	**-2.23**	**-.44**	**-2.31***	**[-4,18, -0.29]**	

Note. *N* = 52. The first step included two control variables: Age and Education, and Age was excluded because of its weak predicting power for the model. Group: Treatment Group (Waiting-List = 0, CBM-I = 1). T1-WSAP-H: pre-test Word Sentence Association Paradigm-Hostile Interpretation; △AQCV-PA: change of physical aggression score (post-test AQCV-PA minus pre-test AQCV-PA). ***p < .001, **p < .01, *p < .05.

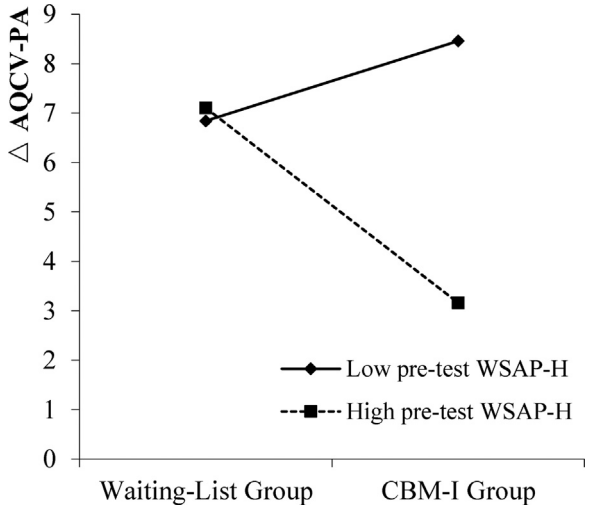


Figure 2 Interaction between Group and Initial WSAP-HS Score in Predicting △AQCV-PA.
Note. *n* for Waiting-List group = 25; *n* for CBM-I group = 27; WSAP-H = Word Sentence Association Paradigm-Hostile Interpretation. For participants with a low level of pre-test WSAP-H, the △AQCV-PA in the CBM-I group was not significantly different from that of the Waiting-List group (*simple slope* = 1.62, *t* = 0.96, *p* = .342); for participants with a high level of pre-test WSAP-H, the △AQCV-PA in the CBM-I group was significantly smaller than that of the Waiting-List group (*simple slope* = -3.95, *t* = -2.32, *p* = .024).

positive reinforcement for juvenile delinquents' positive interpretations helped them to interpret ambiguous social cues more positively.

The intervention effect of CBM-I on hostile interpretation was not significant, which may be affected by the characteristics of juvenile delinquents and the CBM-I tasks. Firstly, the impulsivity, irritability, lower punishment sensitivity, and existing hostile beliefs (Guan et al., 2015; Morgan et al., 2014) might impair the juvenile delinquents' readiness for intervention, elicit ambivalent reactions and greater treatment resistance and thus impede their engagement in therapy, which thereby can diminish therapeutic change (Howells & Day, 2003). Secondly, the CBM-I tasks, during which different feedback was given for positive and hostile interpretations of ambiguous situations, might be helpful to encourage juvenile delinquents to adopt positive interpretations, but less effective in helping them reject hostile interpretations (Gonsalves et al., 2019). Presenting the hostile interpretation option may somehow prompt and amplify the hostile cues that may be implied in ambiguous situations, and therefore might activate the hostile schemas of juvenile delinquents, leading them to expect to be threatened or provoked and thus, think in hostile ways (Gilbert & Daffern, 2010). Lastly, there were only four treatment sessions, the relatively short-term interventions might have a limited effect on reducing hostile interpretation bias (Menne-Lothmann et al., 2014).

The intervention effects on HIBT and AIHQ were not significant, which suggests that CBM-I as used in this study increased a positive interpretation bias towards ambiguous scenario, but did not reduce a hostile interpretation bias toward others' intentions and facial expressions. We believe that differences between the intervention materials and some of the outcome measures may have affected these results (Gonsalves et al., 2019; Hertel & Mathews, 2011). For example, the CBM-I materials were quite different from the HIBT task employed to measure bias in interpreting facial expressions. Indeed, related empirical studies (LeMoult et al., 2018; Podina, Cosmoiu, Rusu, & Chivu, 2020; Salemink, Woud, Roos, Wiers, & Lindgren, 2019) and meta-analysis research (Gonsalves et al., 2019; Hertel & Mathews, 2011) have reported that inconsistency between the intervention materials and outcome indicators can reduce the estimated treatment effect and raise concerns about whether CBM-I generalizable to other materials and situations. Another possible explanation for the results is that HIBT and AIHQ did not include a positive condition and failed to provide estimations of changes in positive interpretation bias (Chen et al., 2012; Smeijers et al., 2017). Positive interpretations and negative/hostile interpretations are on two distinct continuums, instead of the opposite ends of a single continuum (Beard & Amir, 2009; Dillon et al., 2016; Huppert et al., 2003), and the present study provides supportive evidence regarding the distinctiveness of the intervention effects on positive interpretation and hostile interpretation.

It is worth noting that hostile interpretation bias and positive interpretation bias were also reduced significantly among the Waiting-List group, which suggests the existence of confounding factors. Specifically, the reduction of the hostile interpretation bias might be related to daily educational programs in the juvenile correctional institution. In China's judicial system, educational programs, such as social skills and interpersonal communication training, are used to guide the discipline of juvenile delinquents. Previous studies have revealed that education programs regarding social interaction skills could reduce the level of hostile interpretation bias in juvenile delinquents (van der Stouwe et al., 2016). Although the Waiting-List group did not receive CBM-I, the daily education programs may have reduced their interpretation bias. Participants in the CBM-I group also attended these educational programs, but as mentioned above, the beneficial effect of which may offset by the hostile interpretation options in the intervention program that might reinforce their hostile interpretation bias (Gilbert & Daffern, 2010). However, the reason for the reduction of positive interpretation bias in the Waiting-List group is not clear. A combination of individual, situational, and organizational factors might have produced a change in the Waiting-List group (Howells & Day, 2003). This should be investigated in further research on involvement in more than one program at a time.

The moderation analysis showed that the intervention effect of CBM-I on self-reported aggressive behavior was affected by the participants' initial hostile interpretation bias. Only for the juvenile delinquents with high initial hostile interpretation scores, the intervention effect on self-reported physical aggressive behavior was significant. One explanation may be the limited variability in the hostile interpretation bias scores. According to the premises of CBM-I, the decline in aggression occurs as a function of decreases in the level of hostile interpretation bias (Jones & Sharpe, 2017), and for individuals with low initial hostile interpretation scores, there is not much room to improve (Gonsalves et al., 2019). This may limit the intervention effect of CBM-I on the self-reported aggression of youth with relatively low hostile interpretation bias.

Furthermore, we found that CBM-I reduced self-reported physically aggressive behavior but not the other forms of aggression, namely self-reported verbal aggression, angry, hostile and self-directed aggression. This pattern might be explained by the characteristics of the CBM-I intervention and of the participants. Firstly, there is evidence that cognitive therapy, which is the basis of CBM-I, is more effective for physically aggressive behavior than for other types of aggression (Cristea et al., 2015; Menne-Lothmann et al., 2014; Ross et al., 2013). Secondly, the scores for physical aggression in the current sample were higher than the scores for the other types of aggression, which allowed for greater improvement in treatment. Moreover, physically aggressive behavior may be more easily detected and thus more often punished in juvenile correctional institutions. Therefore, juvenile delinquents may be more motived to reduce their physical aggressive behavior, and may be more amenable to treatment (Howells & Day, 2003). For these reasons, we conclude that in this population, CBM-I is more likely to have an impact on self-reported physical aggression than on other forms of aggression.

The present study has some limitations. The sample size was relatively small, with only 52 valid participants. The small sample was related to heterogeneity on the pre-test and post-test scores, and this variability might limit the generalization of the results (Well et al., 1990). It is necessary to increase the sample size in future studies. In addition, male juvenile delinquents were the participants in this study. Male and female juvenile delinquents have been shown to differ in the cognitive and emotional processing of anger, and they might benefit from different training programs (Suter et al., 2002). Indeed, whether the current results can be extrapolated to the sample of female juvenile delinquents needs to be addressed in future research. The study was conducted within a juvenile correctional institution and participants in the treatment group and the Waiting-List group may share the information about the intervention. To avoid the confounding effect, a placebo-group should be included in further studies. Another limitation is what aggressive behavior was estimated with self-reported measures, which might be affected by response bias and social desirability. A tendency toward deception has been reported in the forensic population (Mayorga-Sierra, Novo, Fariña, & Seijo, 2020; Novo, Fariña, Seijo, & Arce, 2012). Future researchers are encouraged to use multi-method assessment to circumvent subjectivity and mono-method biases. Considering the characteristics of juvenile delinquents and the particularity of the environment of the juvenile correctional institution, there may be some confounding variables that affect the current results, such as the total detention time imposed on different youth, the detention time participants had met, whether participants had previously carried out other educational programs, the supervision orders or parole conditions of the institution and the institution's educational programs. In the future, researchers need to fully consider the roles of these possible confounding variables.

This present findings have implications for future research and clinical practice. Firstly, the moderation analyses indicated that the intervention effect differed across participants, suggesting the importance of selecting or matching participants based on the initial level of hostile interpretation bias (Cantos et al., 2019; Jones & Sharpe, 2017). It is necessary to further explore the influence of individual characteristics on the intervention effect and conduct corresponding intervention-population matching to maximize the intervention effect and saving resources. In clinical practice, the intervention tasks, materials, and presentation of materials should also be evaluated. As discussed above, for juvenile delinquents, tasks that provide the option to make a hostile interpretation about a stimulus may to some extent inadvertently maintain their hostile interpretation mode (Gilbert & Daffern, 2010). In the future, researchers must compare the intervention effects of different CBM-I programs and combine them to harness the strengths of different approaches (Gonsalves et al., 2019). Thirdly, consistent with previous research, the present study also found a limited generalization of intervention effects (LeMoult et al., 2018; Podina et al., 2020; Salemink et al., 2019). The effects of CBM-I based on specific materials (such as ambiguous situations) may be difficult to generalize to another materials (such as facial expressions), which limit the large-scale application of CBM-I in clinical settings. It is necessary for future research to further explore the factors that affect

the generalization of the effect of CBM-I intervention, and try to improve the ecological validity of this treatment.

Finally, in addition to paying attention to the cognitive factors that trigger aggression, it is necessary to consider the roles of broader physiological-psychological-social factors related to aggression, especially the criminogenic and non-criminogenic needs of offenders (Bonta & Andrews, 2017). On the one hand, intervention should target criminogenic needs (e.g., antisocial cognitions and other factors that increase the risk of recidivism) of juvenile offenders. On the other hand, non-criminogenic needs, including personal, social, and psychological needs that are not necessarily associated with recidivism can be tested as protective factors in this population (Arias et al., 2020). Aside from delivering CBM-I, future intervention programs should also meet the non-criminogenic needs of juvenile delinquents, such as improving psychological adjustment and social skills.

Funding and acknowledge

This work was funded by The National Social Science Fund of China (16CSH051). The funding body played no role in study design, data collection, analysis, interpretation, or in writing the manuscript. We are grateful to the juvenile delinquents for their volunteering participation, to the staff of the juvenile correctional institution for their kind support. We also want to thank Lizu Lai, Yuzhong Lin, Huixian Lin, Siqi Tan, and Linlin Shi for their efforts in data collection and preparation for the research.

References

Almoghrabi, N., Huijding, J., & Franken, I. H. (2018). The effects of a novel hostile interpretation bias modification paradigm on hostile interpretations, mood, and aggressive behavior. *Journal of Behavior Therapy and Experimental Psychiatry*, *58*, 36–42. https://doi.org/10.1016/j.jbtep.2017.08.003

Arias, E., Arce, R., Vázquez, M. J., & Marcos, V. (2020). Treatment efficacy on the cognitive competence of convicted intimate partner violence offenders. *Annals of Psychology*, *36*, 427–435. https://doi.org/10.6018/analesps.428771

Assink, M., van der Put, C. E., Hoeve, M., de Vries, S. L., Stams, G. J. J., & Oort, F. J. (2015). Risk factors for persistent delinquent behavior among juveniles: A meta-analytic review. *Clinical Psychology Review*, *42*, 47–61. https://doi.org/10.1016/j.cpr.2015.08.002

Beard, C., & Amir, N. (2009). Interpretation in social anxiety: When meaning precedes ambiguity. *Cognitive Therapy and Research*, *33*, 406–415. https://doi.org/10.1007/s10608-009-9235-0

Beard, C., Rifkin, L. S., Silverman, A. L., & Björgvinsson, T. (2019). Translating CBM-I into real-world settings: Augmenting a CBT-based psychiatric hospital program. *Behavior Therapy*, *50*, 515–530. https://doi.org/10.1016/j.beth.2018.09.002

Bonta, J., & Andrews, D. A. (2017). *The psychology of criminal conduct* (6th ed.). Routledge.

Brettschneider, M., Neumann, P., Berger, T., Renneberg, B., & Boettcher, J. (2015). Internet-based interpretation bias modification for social anxiety: A pilot study. *Journal of Behavior Therapy and Experimental Psychiatry*, *49*, 21–29. https://doi.org/10.1016/j.jbtep.2015.04.008

Cantos, A. L., Kosson, D. S., Goldstein, D. A., & O'Leary, K. D. (2019). Treatment impact on recidivism of family only vs. generally violent partner violence perpetrators. *International Journal of Clinical and Health Psychology*, *19*, 171–180. https://doi.org/10.1016/j.ijchp.2019.05.002

Chen, X. Q., Dong, Y., Dennis, R. C., Wang, K., Zhong, H., Xia, H. S., Kong, X. M., Geng, F., Fang, P., Chu, Z. X., Ji, F., Chen, J., & Xie, S. S. (2012). Reliability and validity of the Chinese version of Ambiguous Intentions Hostility Questionnaire. *Chinese Journal of Behavioral Medicine and Brain Science*, *21*, 1038–1104.

Combs, D. R., Penn, D. L., Wicher, M., & Waldheter, E. (2007). The Ambiguous Intentions Hostility Questionnaire (AIHQ): A new measure for evaluating hostile social-cognitive biases in paranoia. *Cognitive Neuropsychiatry*, *12*, 128–143. https://doi.org/10.1080/13546800600787854

Crick, N. R., & Dodge, K. A. (1994). A review and reformulation of social information–processing mechanisms in children's social adjustment. *Psychological Bulletin*, *115*, 74–101.

Cristea, I. A., Mogoase, C., David, D., & Cuijpers, P. (2015). Practitioner Review: Cognitive bias modification for mental health problems in children and adolescents: A meta-analysis. *Journal of Child Psychology and Psychiatry*, *56*, 723–734. https://doi.org/10.1111/jcpp.12383

Dietel, F. A., Zache, C., Bürkner, P. C., Schulte, J., Möbius, M., Bischof, A., ... & Buhlmann, U. (2020). Internet-based interpretation bias modification for body dissatisfaction: A three-armed randomized controlled trial. *International Journal of Eating Disorders*, *53*, 972–986.

Dillon, K. H., Allan, N. P., Cougle, J. R., & Fincham, F. D. (2016). Measuring hostile interpretation bias: The WSAP-hostility scale. *Assessment*, *23*, 707–719. https://doi.org/10.1177/1073191115599052

Gilbert, F., & Daffern, M. (2010). Integrating contemporary aggression theory with violent offender treatment: How thoroughly do interventions target violent behavior? *Aggression and Violent Behavior*, *15*, 167–180. https://doi.org/10.1016/j.avb.2009.11.003

Gonsalves, M., Whittles, R. L., Weisberg, R. B., & Beard, C. (2019). A systematic review of the word sentence association paradigm (WSAP). *Journal of Behavior Therapy and Experimental Psychiatry*, *64*, 133–148. https://doi.org/10.1016/j.jbtep.2019.04.003

Guan, M., Liao, Y., Ren, H., Wang, X., Yang, Q., Liu, X., & Wang, W. (2015). Impaired response inhibition in juvenile delinquents with antisocial personality characteristics: A preliminary ERP study in a Go/Nogo task. *Neuroscience Letters*, *603*, 1–5. https://doi.org/10.1016/j.neulet.2015.06.062

Hawkins, K. A., & Cougle, J. R. (2013). Effects of interpretation training on hostile attribution bias and reactivity to interpersonal insult. *Behavior Therapy*, *44*, 479–488. https://doi.org/10.1016/j.beth.2013.04.005

Hertel, P. T., & Mathews, A. (2011). Cognitive bias modification: Past perspectives, current findings, and future applications. *Perspectives on Psychological Science*, *6*, 521–536. https://doi.org/10.1177/1745691611421205

Howells, K., & Day, A. (2003). Readiness for anger management: Clinical and theoretical issues. *Clinical Psychology Review*, *23*, 319–337. https://doi.org/10.1016/S0272-7358(02)00228-3

Huppert, J. D., Foa, E. B., Furr, J. M., Filip, J. C., & Mathews, A. (2003). Interpretation bias in social anxiety: A dimensional perspective. *Cognitive Therapy and Research*, *27*, 569–577. https://doi.org/10.1023/A:1026359105456

Jones, E. B., & Sharpe, L. (2017). Cognitive bias modification: A review of meta-analyses. *Journal of Affective Disorders*, *223*, 175–183. https://doi.org/10.1016/j.jad.2017.07.034

Koster, E. H. W., Fox, E., & MacLeod, C. (2009). Introduction to the special section on cognitive bias modification in emotional disorders. *Journal of Abnormal Psychology*, *118*, 1–4. https://doi.org/10.1037/a0014379

Krug, E. G., Mercy, J. A., Dahlberg, L. L., & Zwi, A. B. (2002). The world report on violence and health. *The Lancet*, *360*, 1083–1088. https://doi.org/10.1016/S0140-6736(02)11133-0

LeMoult, J., Colich, N., Joormann, J., Singh, M. K., Eggleston, C., & Gotlib, I. H. (2018). Interpretation bias training in depressed adolescents: Near-and far-transfer effects. *Journal of Abnormal Child Psychology*, *46*, 159–167. https://doi.org/10.1007/s10802-017-0285-6

Li, X. Y., Fei, L. P., Zhang, Y. L., Nui, Y. J., Tong, Y. S., & Yang, S. J. (2011). Reliability and validity of the Chinese version of Buss & Perry Aggression Questionnaire. *Chinese Journal of Nervous and Mental Diseases*, *37*, 607–613.

Liu, J., Zhou, Y., & Wenyu, G. U. (2009). Reliability and validity of Chinese version of Buss-Perry Aggression Questionnaire in adolescents. *Chinese Journal of Clinical Psychology*, *17*, 449–451.

Lochman, J. E., & Dodge, K. A. (1994). Social-cognitive processes of severly violent, moderately aggressive, and nonaggressive boys. *Journal of Consulting and Clinical Psychology*, *62*, 366–374. https://doi.org/10.1037/0022-006X.62.2.366

Lu, Q., Guo, K. Y., Zhang, M., Zhang, X. B., Hu, F. Q., & Yang, J. L. (2018). Chinese Juvenile Delinquency Research Report 2017: Based on a Comparative Study of Juvenile Offenders and Other Groups. *Juvenile Delinquency*, *219*, 31–45.

MacDonald, E. M., Koerner, N., Antony, M. M., Vickers, K., Mastorakos, T., & Kuo, J. (2020). Investigating the therapeutic potential of cognitive bias modification for high anxiety sensitivity. *Journal of Behavior Therapy and Experimental Psychiatry*, *68*, Article 101521 https://doi.org/10.1016/j.jbtep.2019.101521

Matheson, E., Wade, T. D., & Yiend, J. (2019). Utilizing cognitive bias modification to remedy appearance and self-worth biases in eating disorder psychopathology: A systematic review. *Journal of Behavior Therapy and Experimental Psychiatry*, *65*, Article 101482 https://doi.org/10.1016/j.jbtep.2019.101482

Mayorga-Sierra, É., Novo, M., Fariña, F., & Seijo, D. (2020). Needs analysis for the personal, social, and psychological adjustment of adolescents at risk of delinquency and juvenile offenders. *Anales de Psicología/Annals of Psychology*, *36*, 400–407. https://doi.org/10.6018/analesps.428631

Menne-Lothmann, C., Viechtbauer, W., Höhn, P., Kasanova, Z., Haller, S. P., Drukker, M., van Os, J., Wichers, M., & Lau, J. Y. (2014). How to boost positive interpretations? A meta-analysis of the effectiveness of cognitive bias modification for interpretation. *Plos One*, *9*, Article e100925 https://doi.org/10.1371/journal.pone.0100925

Micco, J. A., Henin, A., & Hirshfeld-Becker, D. R. (2014). Efficacy of interpretation bias modification in depressed adolescents and young adults. *Cognitive Therapy and Research*, *38*, 89–102. https://doi.org/10.1007/s10608-013-9578-4

Morgan, J. E., Bowen, K. L., Moore, S. C., & Van Goozen, S. H. (2014). The relationship between reward and punishment sensitivity and antisocial behavior in male adolescents. *Personality and Individual Differences*, *63*, 122–127. https://doi.org/10.1016/j.paid.2014.01.054

Novo, M., Fariña, F., Seijo, M. D., & Arce, R. (2012). Assessment of a community rehabilitation programme in convicted male intimate-partner violence offenders. *International Journal of Clinical and Health Psychology*, *12*, 219–234.

Podina, I. R., Cosmoiu, A., Rusu, P., & Chivu, A. (2020). Positive thinking is not adaptive thinking: A cognitive-behavioral take on Interpretation Bias Modification for social anxiety. *Journal of Rational-Emotive & Cognitive-Behavior Therapy*, *38*, 424–444. https://doi.org/10.1007/s10942-020-00344-5

Ross, J., Quayle, E., Newman, E., & Tansey, L. (2013). The impact of psychological therapies on violent behaviour in clinical and forensic settings: A systematic review. *Aggression and Violent Behavior*, *18*, 761–773. https://doi.org/10.1016/j.avb.2013.09.001

Salemink, E., & Wiers, R. W. (2011). Modifying threat-related interpretive bias in adolescents. *Journal of Abnormal Child Psychology*, *39*, 967–976. https://doi.org/10.1007/s10802-011-9523-5

Salemink, E., Woud, M. L., Roos, M., Wiers, R., & Lindgren, K. P. (2019). Reducing alcohol-related interpretive bias in negative affect situations: Using a scenario-based Cognitive Bias Modification training paradigm. *Addictive Behaviors*, *88*, 106–113. https://doi.org/10.1016/j.addbeh.2018.07.023

Smeijers, D., Rinck, M., Bulten, E., van den Heuvel, T., & Verkes, R. J. (2017). Generalized hostile interpretation bias regarding facial expressions: Characteristic of pathological aggressive behavior. *Aggressive Behavior*, *43*, 386–397. https://doi.org/10.1002/ab.21697

Suter, J., Byrne, M., Byrne, S., Howells, K., & Day, A. (2002). Anger in prisoners: Women are different from men. *Personality and Individual Differences*, *32*, 1087–1100. https://doi.org/10.1016/S0191-8869(01)00105-2

Tuente, S. K., Bogaerts, S., & Veling, W. (2019). Hostile attribution bias and aggression in adults-a systematic review. *Aggression and Violent Behavior*, *46*, 66–81. https://doi.org/10.1016/j.avb.2019.01.009

van der Stouwe, T., Asscher, J. J., Hoeve, M., van der Laan, P. H., & Stams, G. J. J. (2016). Social skills training for juvenile delinquents: Post-treatment changes. *Journal of Experimental Criminology*, *12*, 515–536. https://doi.org/10.1007/s11292-016-9262-2

Vassilopoulos, S. P., Brouzos, A., & Andreou, E. (2015). A multi-session attribution modification program for children with aggressive behavior: Changes in attributions, emotional reaction estimates, and self-reported aggression. *Behavioural and Cognitive Psychotherapy*, *43*, 538–548. https://doi.org/10.1017/S1352465814000149

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

Well, A. D., Pollatsek, A., & Boyce, S. J. (1990). Understanding the effects of sample size on the variability of the mean. *Organizational Behavior and Human Decision Processes*, *47*, 289–312. https://doi.org/10.1016/0749-5978(90)90040-G

Wilkowski, B. M., & Robinson, M. D. (2008). The cognitive basis of trait anger and reactive aggression: An integrative analysis. *Personality and Social Psychology Review*, *12*, 3–21. https://doi.org/10.1177/1088868307309874

Yeung, E. S., & Sharpe, L. (2019). Cognitive bias modification for social anxiety: The differential impact of modifying attentional and/or interpretation bias. *Cognitive Therapy and Research*, *43*, 781–791. https://doi.org/10.1007/s10608-019-10012-3