Development and Validation of the Emotion Regulation Ability Test for Chinese Youth

Yi Ming Li†‡, Jian Li†, Hong Zou† and Shengnan Wei†

Abstract
A culture- and age-appropriate instrument for measuring emotion regulation ability is needed for the research and practice of Chinese adolescents’ emotion regulation. This study developed and validated a situational judgment test of emotion regulation ability for Chinese youth (STER-CY). Three samples were recruited, and approximately 4380 5th- to 11th-grade students (but no 9th-grade students) participated in the study. Researchers collected emotional situations and responses based on the life of indigenous samples and examined the reliability and validity of the test scores. The results showed that Cronbach's alpha and test–retest correlations provided evidence for the reliability of the test scores. Exploratory and confirmatory factor analysis supported unidimensionality. Construct validity was further verified by convergent and discriminant validity. Criteria-related validity was confirmed by the correlations between this test and some outcome variables related to emotion regulation. It was also found that girls scored higher on this test than boys did and that emotion regulation ability significantly increased from 5th to 7th grade, but it did not improve from 7th to 11th grade. Considered together, these findings showed that the STER-CY is a psychometrically sound measure of emotion regulation ability and can be used in future research and practice.

During adolescence, children experience more frequent labile and intense emotions compared to their younger years (Riediger, Schmieck, Wagner, & Lindenberger, 2009). The ability to regulate their emotions partly determines whether adolescents can complete the developmental tasks of the adolescence period (McLaughlin, Hatzenbuehler, Mennin, & Nolen-Hoeksema, 2011). Specifically, emotion regulation ability is positively related to social competence, prosocial behaviors, academic achievements and peer relationships, and negatively related to emotional and behavioral problems (e.g., Riediger & Klipker, 2014; Silk, Steinberg, & Sheffield, 2003; Zeman, Cassano, Ferry-Parrish, & Stegall, 2006).

Considering the importance of emotion regulation ability for adolescence, the development of a measurement instrument has significant theoretical and practical implications. Because cultural differences (e.g., Miyamoto, Ma, & Petermann, 2014) and age differences (e.g., Zimmermann & Iwanski, 2014) exist in the field of emotion regulation, an age-appropriate indigenous instrument is needed for research on the emotion regulation ability of Chinese adolescents. The purpose of the current study is to develop and validate such an instrument. Before elaborating on the development of the instrument, we will first discuss the definition and measurement of, and cultural and age differences in emotion regulation ability.

Defining emotion regulation ability
Emotion regulation refers to the processes of initiating, maintaining or modulating the occurrence, form, intensity or duration of emotional reactions to accomplish one’s goals (Eisenberg & Sparind, 2004). According to the definition, three core elements are involved in the process of emotion regulation (Gross, 2014), including goals, strategies and outcomes of emotion regulation. Strategy use and regulation outcomes have been used to evaluate emotion regulation ability (Mayer, Salovey, Caruso, & Sitarenios, 2003; Gratzi & Roemer, 2004). However, emotional regulation goals might not be a good indicator. Specifically, emotional goals are influenced by many factors and might be different for people with the same level of emotion regulation ability in similar situations. For example, European Americans may show their excitement wholeheartedly after winning a contest, while East Asians might restrain themselves from expressing their enthusiasm and display humility. In this case, an individual’s emotional goals are determined by display rules that are shaped by culture (Engelmann & Pogosyan, 2013). In America, a high arousal of positive emotion is desired, but East Asian cultures prefer a low arousal of positive emotion (Tsai, Knutson, & Fung, 2006). Similarly, emotional goals could be influenced...
Lopes, Salovey, and Straus (2003) found that the MSCEIT scores on SJTs might weaken the influence of socially desirable responses. As far as we know, most of the popular emotion regulation scales, such as the Mayer-Salovey-Caruso Emotional Intelligence Test (MSCEIT; Mayer et al., 2003) and the Emotion Regulation Questionnaire (ERQ; Gross & John, 2003), focus on self-regulation of emotion. The exceptions are the emotion management subscale of the Mayer-Salovey-Caruso Emotional Intelligence Test – Youth Version (MSCET-YV; Rivers, Brackett, Reyes, Mayer, Caruso, & Salovey, 2012) and the Emotion Regulation of Others and Self (EROS) scale (Niven, Totterdell, Stride, & Holman, 2011). Not only do they include the items of both intrinsic and extrinsic regulation, but Niven et al. (2011) found that intrinsic regulation was correlated with extrinsic regulation in terms of strategy use. Thus, some research evidence supports that emotion regulation ability might involve the skills to regulate both one’s own emotions and others’ emotions.

Taken together, emotion regulation ability could be defined as a skill to change one’s own and others’ emotions by using effective strategies to achieve a desired emotion goal. Next, we will discuss the assessment of emotion regulation ability.

Measuring emotion regulation ability

Extant instruments mainly take the form of self-report rating scales and situational judgment tests (SJTs). Self-report rating scales can provide plenty of information about individuals’ typical or habitual response to emotional events and emotion regulation outcomes. However, self-report ratings have limitations that might be hard to overcome. First, self-report ratings of emotion regulation ability are susceptible to socially desirable responses. Individuals might evaluate their emotion regulation ability in a positive way that meets social expectations because of the desirable nature of emotion ability and transparent content of self-report scale items (Day & Carroll, 2008; Keefe, 2015). Second, the accuracy of self-report ratings largely depends on self-awareness. A study by Sheldon, Dunning, and Ames (2014) showed that individuals with the lowest level of emotional competencies were most likely to overestimate their own emotional competencies. In this case, deficits in self-awareness might cause inaccuracy in self-report ratings. Third, many self-report rating scales of emotion regulation ability ignore the context in which emotions need to be regulated. The same strategy could cause different results in different situations (Butler, Lee, & Gross, 2007; Shepess, Catran, & Meiran, 2009). Without knowing the demands of a particular situation, it is hard to evaluate whether people are using the best strategy. Therefore, self-report rating scales might not provide adequate and accurate evidence to evaluate an individual’s emotion regulation ability.

SJTs are designed to measure individuals’ judgment concerning possible responses to a situation and are usually used in work-related situations (McDaniel & Nguyen, 2001). Some researchers developed SJTs of emotional competencies (e.g., the MSCEIT) considering the following advantages. First, SJTs could overcome the limitations of self-report rating scales to some extent. Specifically, SJTs might weaken the influence of socially desirable responses. Lopes, Salovey, and Straus (2003) found that the MSCEIT scores were not associated with social desirability. Moreover, SJTs do not need people to rate their own abilities and could avoid the constraint of self-awareness. Second, SJTs provide the context in which emotion and emotion regulation occur. They measure not only the strategy repertoire that people have but also whether they flexibly select an appropriate strategy to address a situation. Therefore, SJTs provide more information about individuals’ emotion regulation ability.

The main limitation of SJTs is that they are not able to measure how people conduct regulation behaviors but instead measure knowledge about regulation strategy. However, a meta-analysis of organizational behavior studies has showed that SJT scores could well predict people’s job performance (McDaniel, Morgeson, Finnegan, Campion, & Braverman, 2001). It might suggest that the results of SJTs can be used to predict how well individuals implement regulation strategies in a specific situation. Therefore, when balancing the positives and negatives, SJTs might be a better way to assess emotion regulation ability than self-report rating scales.

An available SJT for measuring adolescents’ emotion regulation ability is the emotion management subscale of the MSCEIT-YV (Rivers et al., 2012). The test items in the MSCEIT-YV cover intrinsic and extrinsic regulation. Specific emotional goals are included in the situations and serve as a part of the question stems. Moreover, the goals involve not only hedonic regulation (e.g., alleviating negative emotions and generating positive emotions) but also instrumental goals (e.g., weakening positive emotions and triggering negative emotions for some goals; Rivers et al., 2012). Thus, having the same goals of emotion regulation in the same situation is more likely to ensure that participants’ choices reflect their ability to achieve the goal rather than their culture differences or personal beliefs. Therefore, compared to other tests, the MSCEIT-YV provides a more comprehensive and accurate assessment of emotion regulation ability. However, there are only six situations in this test, which means that the representativeness of emotional situations might be limited. Additionally, we cannot use the MSCEIT-YV to assess Chinese adolescents’ emotion regulation ability because of possible cultural differences.

Cultural and age differences in emotion regulation ability

Extant studies have provided evidence that emotion regulation could vary across cultures (Butler et al., 2007) and different age groups (Zimmermann & Iwanski, 2014). Cultural and age differences need to be considered when developing an emotion regulation instrument.

Cultural differences can be found in emotion regulation goals and strategies. First, for emotional goals, people from Eastern countries are less motivated than people from Western countries to engage in hedonic emotion regulation after negative events (e.g., enhance positive emotions and reduce negative emotions) because dialectical beliefs that happiness and sadness are connected prevail in Eastern cultures (Miyamoto et al., 2014). In addition, American individuals value high-arousal positive emotions (e.g., excitement), which can contribute to influencing others, while Chinese individuals value low-arousal positive emotions (e.g., calm), which are good for interpersonal harmony (Tsai et al., 2006). This helps to explain why Chinese people are more likely to set an emotional goal of downregulation (i.e., decrease and minimize the intensity of emotion experience; Sang, Deng, & Luan, 2014). Research findings showed that downregulation could be beneficial to Chinese social adaptation (Sang et al., 2014; Tsai et al., 2006). Second, for emotion regulation strategies, people with Asian values use emotional suppression strategies more frequently.
than people with Western values (Butler et al., 2007). Moreover, the possible negative effects of emotional suppression are more likely to occur for people from Western countries than for Chinese people (Soto, Perez, Kim, Lee, & Minnick, 2011).

Age-related changes in emotion regulation ability exhibit a general trend of increasingly adaptive emotion regulation with increasing age (John & Gross, 2004; Zimmermann & Iwanski, 2014). First, older adolescents use adaptive regulation strategies more and maladaptive strategies less than younger adolescents. For example, from late childhood to adolescence, individuals are more likely to use explanation and negotiation strategies in anger regulation among friends instead of maladaptive strategies such as confrontation, hurt and neglect (Salisch & Vogelgesang, 2005). Second, individuals have more types of and more complex strategies (Zimmermann & Iwanski, 2014). For example, in addition to behavioral distraction, more cognitively distracting attention strategies are used with an increase in age (Zimmer-Gembeck & Skinner, 2011). Third, individuals become better at choosing relatively effective strategies according to situational demands. For example, adolescents are more likely to use problem-solving strategies to address solvable problems, such as issues related to studies and sports, and use distracting attention strategies to handle uncontrollable stress, such as parents’ diseases (Zimmer-Gembeck & Skinner, 2011).

However, the development of emotion regulation ability might not be linear. Zhang (2013) found that 7th-grade students scored higher in emotion regulation ability tests than 8th-, 10th- and 11th-grade students. Additionally, Zimmer-Gembeck and Skinner (2011) found that the use of maladaptive strategies such as cognitive escape and rumination could increase, but not decrease, during a period in adolescence. More studies are needed to examine how emotion regulation ability develops with age.

Considering cultural and age differences in emotion regulation, we cannot use the instruments that have been developed by using samples of Western individuals or adults to evaluate Chinese adolescents’ emotion regulation ability. It is necessary to design a culture- and age-appropriate instrument for research on the emotion regulation ability of Chinese adolescents.

The present study

The present study aimed to develop a situational test of emotion regulation ability for Chinese youth (STER-CY) aged from 11 to 17 years. Similar to the format of the MSCEIT-YV emotion management subscale, each item of this test was composed of a situation with an emotional goal and four response options (see Appendix). According to prior research (e.g., Niven, Totterdell, & Holman, 2009; Rivers et al., 2012), emotional goals could be to regulate one’s own emotions, including recovering from negative emotions, maintaining positive emotions, and generating a particular emotion. Emotional goals could also be to regulate others’ emotions, including improving and worsening others’ emotions. The current study examined the reliability and validity of the STER-CY among 5th- to 11th-grade students. Validity evidence was based on: (a) the results of exploratory and confirmatory factor analysis, (b) the correlations between the STER-CY and other tests related to emotion regulation ability (convergent validity), (c) the correlations between the STER-CY and personality tests (discriminant validity), and (d) the correlations between the STER-CY and some outcome variables of emotion regulation (criteria-related validity). Specifically, the criteria included three kinds of factors – emotional status, interpersonal relationships, and school accommodation – that have been proven to be associated with emotion regulation (e.g., McLaughlin et al., 2011; Zeman et al., 2006). The current study also used the STER-CY to examine gender and grade differences in emotion regulation ability of Chinese adolescents.

Method

Item generation

According to the common procedure for constructing SJTs (McDaniel & Nguyen, 2001), the development of the test items includes three stages: generation of situations, generation of response options, and expert scoring.

Generation of situations

The first step is to screen the typical situations that require emotion regulation for adolescents. First, focus group interviews were conducted to collect emotional situations from 56 5th- to 11th-grade students (28 females, no 9th-grade students) who were recruited from a public school in Beijing. They were asked to describe emotionally salient events that had happened to their parents, teachers, peers and themselves in the past month. One hundred and seven situations were obtained after deleting similar situations or those in which regulation was not needed. The situations covered intrinsic and extrinsic regulation. Second, 351 5th- to 11th-grade participants (174 females, no 9th-grade students) from another public school in Beijing were recruited to rate the above 107 situations, mainly in terms of the possibility of occurrence and the intensity of emotion that the protagonist felt in that situation. Seventy-two situations were retained in which the levels of the main emotions’ intensity were above 3 (a 5-point Likert-type scale ranging from 1 = not at all to 5 = very strong). At the same time, the possibilities of the situations’ occurrence were above 2 (a 4-point Likert-type scale ranging from 1 = absolutely not to 4 = very likely) across both genders and all grades. These situations were modified into two to four brief sentences and served as the stems of the test items.

Generation of response options

Two hundred and eighty-five students from 5th to 11th grade (137 females, no 9th-grade students) in a Beijing public school were equally divided into three groups. Each group was requested to write the best response and another reasonable response that could help achieve the emotional goals in one-third of 72 situations. The responses that were obviously inappropriate in terms of being hurtful to one’s self and others and not being allowed at school, such as for a physical attack, were deleted. The rest of the responses were summarized into different types for each situation. Nine situation items were removed because the number of response options was no more than 4. Finally, 63 situation items were retained, and each had 5–6 response options.

Expert scoring

First, five Chinese native clinical psychologists with vast experience in adolescents’ emotional health ranked the response options of each item in terms of their effectiveness in a particular situation. The Kendall coefficient of concordance of each item served as an indicator of agreement between the experts, and the items with W < .70 were deleted. The experts discussed the rankings for the rest of items until they reached consensus. Second, the experts rated each response option of each item on a 7-point scale from 1 = very inappropriate to 7 = very appropriate. It is reasonable to...
think that each item has the best response option with the highest mean score of above 5 (moderately appropriate), so the items in which the highest means were lower than 5 were deleted. Thus, 30 items were retained for the next analysis.

The Most-Least Effectiveness Rating Method was adopted in this test. As stated above, the option with the highest mean was used as the best response, while the option with the lowest mean was used as the worst response. For each item, the two options whose means had larger differences from those of the best and worst responses were selected from the rest of the options. Then, each situation item received four response options.

**Participants**

Three samples were recruited in this study. The first sample consisted of 164 7th- to 11th-grade students (86 females; no 9th-grade students; $M_{age} = 14.90, SD = 1.71$) from a public school in Beijing. This sample was used in the item analysis. The second sample included 3471 5th- to 11th-grade students (no 9th-grade students) from 12 public schools in five cities in China. One half of these participants ($n = 1735, 840$ females, 16 missing; $M_{age} = 13.67, SD = 2.23$) were randomly selected for use in the exploratory factor analysis. The other half ($n = 1736, 853$ females, 13 missing; $M_{age} = 13.78, SD = 2.22$) was used in the confirmatory factor analysis (CFA). The third sample, comprising 55 7th- and 8th-grade students (25 females), was recruited to investigate the test–retest reliability.

**Measures**

**STER-CY**

The STER-CY was designed to assess the ability to regulate one’s own and others’ emotions for Chinese adolescents aged 11 to 17 years. The initial questionnaire consisted of 30 items, and the number of items was reduced to 15 after the item analysis. Each item was composed of a situation with an emotion goal and four response options. The emotional goals covered intrinsic regulation (i.e., recovering from negative emotions, maintaining positive emotions, and generating a particular emotion) and extrinsic regulation (i.e., improving and worsening others’ emotions). The participants were asked to identify the best and worst option for the achievement of an emotion goal. For example:

Lily runs for a position in the class cadre. She thought that her classmates would support her, but the result is that she only received a few votes. She feels disappointed and sad. What action would be the most effective to help relieve her disappointment and sadness? What action would be the least effective?

A. Talk about her feelings with her parents or friends.
B. Tell herself that she is the best.
C. Play with her friends.
D. Tell herself: It is not a big deal.

One point was given if the best or worst response was chosen correctly. One point was deducted if the best response was chosen as the worst response, or the worst response was chosen as the best response. Otherwise, the score would be zero. A higher score reflected a higher level of emotion regulation ability.

**MSCEIT-YV**

The emotion management subscale of the MSCEIT-YV (Rivers et al., 2012) was used to measure participants’ ability to regulate their own and others’ emotions. The copyright of the MSCEIT-YV English version is owned by Multi-Health Systems (MHS).

As MHS required, two independent translators translated the MSCEIT-YV from English to Chinese, and then another independent translator who was not familiar with the test completed the translation from Chinese back into English. MHS made a few minor changes to the translated English version, while we made the corresponding changes to the Chinese version. The final version was approved by MHS.

The emotion management subscale of the MSCEIT-YV includes six situations and each has three response options. Participants were required to rate the extent to which the response would help the target character achieve a specific emotion goal on a 5-point scale ($1 = \text{not at all helpful}$ to $5 = \text{very helpful}$). The test scores were calculated by the MHS.

**ERQ**

The Chinese version (Wang, Liu, Li, & Du, 2007) of the 10-item ERQ (Gross & John, 2003) was used to measure habitual expressive suppression and reappraisal. Participants were asked to rate their responses on a 7-point Likert-type scale ranging from $1 = \text{strongly disagree}$ to $7 = \text{strongly agree}$. For the scores of the reappraisal and expressive suppression dimensions in the current study, the reliability coefficients were .81 and .71 respectively.

**Self-control scale**

A 16-item self-control scale (Qu, Zou, & Duan, 2006), revised from the study by Grasmick, Titte, Bursik, and Arneklev (1993), was used to measure the level of self-control. This scale includes three dimensions: impulsivity and risk-seeking, self-center and temper, and simple tendency. Participants were asked to rate their responses on a 5-point Likert-type scale ranging from $1 = \text{strongly disagree}$ to $5 = \text{strongly agree}$. A higher score reflects a lower level of self-control. In the current study, the reliability coefficient was .74.

**Adolescents Personality Scale**

The 50-item Adolescents Personality Scale (Zou, 2003; Zhou, Niu, & Zou, 2000) was used to measure individuals’ personality. The scale includes five dimensions: extraversion, agreeableness, neuroticism, conscientiousness and openness. Participants were asked to rate their responses on a 5-point Likert-type scale ranging from $1 = \text{not like me at all}$ to $5 = \text{like me very much}$. In the current study, the reliability coefficients of the five dimension scores ranged from .83 to .90.

**Friendship Quality Questionnaire**

A revised version (Jin & Zou, 2012) of the Friendship Quality Questionnaire (Parker & Asher, 1993) was used to measure the friendship quality between participants and their best friends. Participants were asked to rate their responses on a 5-point Likert-type scale ranging from $1 = \text{strongly disagree}$ to $5 = \text{strongly agree}$. A higher score reflects a higher level of friendship quality. In the current study, the reliability coefficient was .94.

**Child Behavior Checklist (CBCL)**

The 40-item brief version (Li, Zou, & Wang, 2009) of the CBCL (Achenbach, 1991) was used to measure participants’ school accommodations in emotion and behavior. Participants rated their emotion and behavior in the last six months on a 4-point Likert-type scale ranging from $1 = \text{strongly disagree}$ to $4 = \text{strongly agree}$. A higher score reflects more emotional and behavioral problems. In the current study, the reliability coefficient was .93.
Satisfaction with Peer Relationships Scale
Nine items about satisfaction with friends in the Chinese version (Tian & Liu, 2005) of Huebner’s (1994) Multidimensional Life Satisfaction scale were adopted to assess the quality of participants’ peer relationships. Participants were asked to rate their responses on a 6-point Likert-type scale ranging from 1 = strongly disagree to 6 = strongly agree. A higher score reflects a higher quality of peer relationships. In the current study, the reliability coefficient was .83.

Children’s Depression Inventory (CDI)
The Chinese version (Wu, Lu, Tan, & Yao, 2010) of the CDI (Kovacs & Staff, 2003) was used to measure participants’ depression. Item scores range from 0 (absence of the symptom) to 2 (severe symptom). A higher score reflects a higher level of depression. In the current study, the reliability coefficient was .83.

Children’s Anxiety Scale
The six-item Children’s Anxiety Scale was adapted from the dimension of generalized anxiety disorder of the Spence Children’s Anxiety Scale (SCAS; Spence, 1998) and revised by Zhao et al. (2012). Participants were asked to rate their responses on a 5-point Likert-type scale ranging from 1 = strongly disagree to 5 = strongly agree. A higher score reflects a higher level of anxiety. In the current study, the reliability coefficient was .80.

Procedures
The 30-item STER-CY was administered to the first sample of 164 middle school students in their classroom. An item analysis was conducted, and 15 items were retained in terms of item-total correlations and the representativeness of situation content. The second round of 3471 participants completed the 15-item STER-CY test and a battery of scales were used to test its validity. The participants were divided into five groups and completed a portion of these scales to avoid fatigue and boredom. Specifically, all participants completed the STER-CY. A total of 214 participants completed the emotion management scale of MSCET-YV. A total of 1281 participants completed the Friendship Quality Questionnaire and CBCL. A total of 658 participants completed the Adolescents Personality Scale, the Children’s Anxiety Scale, and ERQ. A total of 757 participants completed the self-control scale, CDI and the Satisfaction with Peer Relationships Scale. The remaining participants only completed the STER-CY.

Results
Item Analysis
Seventeen items with an item-total of r < .30 were removed from the test. However, considering the representativeness and diversity of the situation content, two items from the removed items were put back into the test. Thus, 15 items whose item-total correlations ranged from .24 to .64 were retained and used in the following exploratory factor analysis (EFA).

Exploratory factor analysis and confirmatory factor analysis
An EFA was conducted on the 15-item test to compare the one-factor, two-factor and three-factor solutions with MPlus 7. The maximum likelihood method was used to estimate the parameters. Model fit indices included chi-square, root mean square error of approximation (RMSEA), standardized root mean square residual (SRMR), comparative fit index (CFI), and Tucker-Lewis index (TLI). As presented in Table 1, the fit indices of the three models were all acceptable (McDonald & Ho, 2002). However, for the rotated factor loadings of the items in the two-factor model, only three item loadings on the second factor exceeded .30. Moreover, it was hard to interpret the meaning of each factor. Similarly, for the three-factor model, only one item exceeding .30 loaded on the first factor and two items loaded on the third factor. Considering the goodness of fit and the interpretability of solutions, the one-factor model was the most appropriate for the current data. Consequently, the results of the EFA provided initial evidence for a one-factor structure for the STER-CY.

Another sample was used to perform a CFA to confirm a one-factor construct of the 15-item STER-CY with MPlus 7. The robust maximum likelihood (MLR) estimator was adopted to estimate the parameters. The model fitted well to the data. Specifically, χ² = 274.52, df = 90, χ²/df = 3.05, CFI = .951, TLI = .943, RMSEA = .034, SRMR = .027. The factor loadings of all the items were significant and ranged from .31~.70. The results of the CFA validated the one-factor structure of the STER-CY.

Reliability
Two indices were used to evaluate reliability. First, the internal consistency reliability was assessed using Cronbach’s alpha. The result showed that Cronbach’s alpha was .78. Second, the test-retest reliability was computed with a sample of 55 7th- and 8th-grade students (25 females) and the test was completed twice with an interval of two weeks. The test-retest reliability was .66.

Convergent validity
Convergent validity was tested by correlations between the STER-CY, the emotion management subscale of the MSCET-YV, the ERQ and the self-control scale. The results showed that the STER-CY was positively correlated with the emotion management subscale of the MSCET-YV (r = .24, p < .001) and the reappraisal subscale of the ERQ (r = .16, p < .001). The STER-CY was negatively correlated with self-control (r = −.22, p < .001) and the expressive suppression scale of the ERQ (r = −.09, p = .026).

Discriminant validity
Discriminant validity was tested by correlations between the STER-CY and the five dimensions of personality. The results showed that the STER-CY was negatively correlated with neuroticism (r = −.17, p < .001) and positively correlated with agreeableness (r = .13, p = .001) and conscientiousness (r = .08, p = .037). Additionally, the STER-CY was not correlated with extroversion (r = .02, p = .567) and openness (r = .03, p = .389).

Table 1. Summary of fitness statistics for contrasting alternative models of the STER-CY (n = 1735)

<table>
<thead>
<tr>
<th>Model</th>
<th>χ²</th>
<th>df</th>
<th>RMSEA [90% CI]</th>
<th>SRMR</th>
<th>CFI</th>
<th>TLI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 factor</td>
<td>225.487***</td>
<td>90</td>
<td>.029 [.025, .034]</td>
<td>.025</td>
<td>.965</td>
<td>.959</td>
</tr>
<tr>
<td>2 factor</td>
<td>152.056***</td>
<td>76</td>
<td>.024 [.018, .030]</td>
<td>.020</td>
<td>.980</td>
<td>.973</td>
</tr>
<tr>
<td>3 factor</td>
<td>103.644**</td>
<td>63</td>
<td>.019 [.012, .026]</td>
<td>.016</td>
<td>.989</td>
<td>.982</td>
</tr>
</tbody>
</table>

Note: RMSEA = root mean square error of approximation; CI = confidence interval; SRMR = standardized root mean square residual; CFI = comparative fit index; TLI = Tucker-Lewis index. *p < .01, ***p < .001.
Criteria related validity

Participants’ emotional state, the quality of their peer relationships and school accommodations were used as the criteria. Criteria-related validity was tested by correlations between the STER-CY and the abovementioned criteria. Specifically, the STER-CY was negatively correlated with depression \((r = -.23, p < .001)\), anxiety \((r = -.14, p < .001)\) and the CBCL \((r = -.23, p < .001)\). It was positively correlated with friendship quality \((r = .22, p < .001)\) and satisfaction with peer relationships \((r = .27, p < .001)\).

Table 2 presents the means, standard deviations, reliabilities and correlations between the STER-CY and other variables.

Gender and grade differences

A 5 (grade group: 5th, 6th, 7th, 8th, 10th, 11th) \(\times 2\) (gender) between-subject ANOVA was conducted to examine gender and grade differences in emotion regulation ability. The main effect of gender was significant, \(F(1, 3430) = 55.85\), partial \(\eta^2 = .02\), \(p < .001\). Girls scored significantly higher on the STER-CY than boys. The main effect of grade was significant, \(F(5, 3430) = 3.50\), partial \(\eta^2 = .01\), \(p = .004\). A Bonferroni post hoc test showed that 7th-grade students scored significantly higher than 5th- \((p = .024)\) and 6th-grade students \((p = .013)\). There were no significant differences in scores among 7th- to 11th-grade students. The interactions between gender and grade were not significant.

Discussion

The findings of the current study provide empirical evidence to support the psychometric attributes of the STER-CY, which could serve as a valid measurement instrument for the emotion regulation ability of Chinese adolescents.

Assessment for Emotion Regulation Ability of Chinese Adolescents

The current study operationalized emotion regulation ability as choosing and evaluating an appropriate strategy to regulate one’s own and others’ emotions in intrapersonal and interpersonal situations. Based on this definition, an SJT of emotion regulation ability for 11- to 17-year-old adolescents was developed and validated. The STER-CY has 15 items in which the goals of emotion regulation cover regulating one’s own emotion (i.e., recovering from negative emotions, maintaining positive emotions, and generating a particular emotion), and changing others’ emotions (i.e., improving and worsening others’ emotions). All the situations and responses were derived from the daily life of indigenous samples.

Results from the factor analysis and tests of convergent and discriminant validity provided evidence for the construct validity of the STER-CY. First, the results of the factor analysis supported its unidimensionality. As mentioned above, the MSCEIT-YV includes the items of both intrinsic and extrinsic regulation in its emotion management subscale, but its researchers did not examine the factor structure of emotion regulation ability (Rivers et al., 2012). We went further and found preliminary evidence for the unidimensionality of the MSCEIT-YV. More studies are needed to re-examine the construct of emotion regulation ability and explore the relationship between intrinsic and extrinsic regulation.

Second, the STER-CY was significantly correlated with other tests related to emotion management ability, such as the emotion regulation subscale of the MSCEIT-YV, the ERQ and the self-control scale, suggesting acceptable convergent validity. In particular, the MSCEIT is one of the most popular measure instruments for emotion abilities, so the correlation between the STER-CY and the MSCEIT provided robust evidence for the convergent validity. For the ERQ, reappraisal has been proven to be more adaptive than expressive suppression (Webb, Miles, & Sheeran, 2012). The positive correlation between the STER-CY and reappraisal suggested that individuals with higher emotion regulation ability were more likely to use adaptive strategies.

Last, weak or no correlations between the STER-CY and personality traits supported the discriminant validity. The highest correlation existed between the STER-CY and neuroticism \((r = -.17)\). A possible reason is that the neuroticism items describe negative emotional states, which might reflect the outcome of dysfunctional emotion regulation. Similarly, the second highest correlation was between the STER-CY and agreeableness \((r = .13)\), which might be because the agreeableness items reveal many altruistic attitudes and behaviors that could serve as effective strategies to regulate others’ emotions (Zhou et al., 2000). Therefore, it might be reasonable that the STER-CY was significantly correlated with neuroticism and agreeableness.

For the criteria-related validity, three types of emotion-related variables or outcomes were selected: emotional status (depression and anxiety), interpersonal relationships (quality of relationships with a best friend and classmates), and school accommodation (internal and external problems at school). As expected, all

---

Table 2. Means, standard deviations, internal consistency reliabilities, and correlations between the STER-CY and other variables

<table>
<thead>
<tr>
<th></th>
<th>(N)</th>
<th>(M) (SD)</th>
<th>(\alpha)</th>
<th>(r) Emotion regulation abilities</th>
</tr>
</thead>
<tbody>
<tr>
<td>STER-CY</td>
<td>3471</td>
<td>1.23 (.47)</td>
<td>.78</td>
<td></td>
</tr>
<tr>
<td>Emotion management</td>
<td>214</td>
<td>1.84 (.74)</td>
<td></td>
<td>.24***</td>
</tr>
<tr>
<td>(from MSCEIT-YV)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reappraisal (ERQ)</td>
<td>90</td>
<td>5.00 (1.19)</td>
<td>.81</td>
<td>.16***</td>
</tr>
<tr>
<td>Expressive suppression (ERQ)</td>
<td>90</td>
<td>4.08 (1.48)</td>
<td>.71</td>
<td>-.09*</td>
</tr>
<tr>
<td>Self-control</td>
<td>757</td>
<td>2.80 (.49)</td>
<td>.74</td>
<td>-.22***</td>
</tr>
<tr>
<td>Extraversion</td>
<td>652</td>
<td>3.75 (.80)</td>
<td>.87</td>
<td>.02</td>
</tr>
<tr>
<td>Agreeableness</td>
<td>656</td>
<td>4.07 (.70)</td>
<td>.83</td>
<td>.13**</td>
</tr>
<tr>
<td>Openness</td>
<td>655</td>
<td>3.95 (.72)</td>
<td>.83</td>
<td>.03</td>
</tr>
<tr>
<td>Conscientiousness</td>
<td>653</td>
<td>3.80 (.75)</td>
<td>.90</td>
<td>.08*</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>656</td>
<td>2.53 (.92)</td>
<td>.85</td>
<td>-.17***</td>
</tr>
<tr>
<td>CDI</td>
<td>745</td>
<td>.47 (.38)</td>
<td>.83</td>
<td>-.23***</td>
</tr>
<tr>
<td>Children’s Anxiety scale</td>
<td>651</td>
<td>2.49 (.92)</td>
<td>.80</td>
<td>-.14***</td>
</tr>
<tr>
<td>Friendship quality</td>
<td>1281</td>
<td>3.07 (.66)</td>
<td>.94</td>
<td>-.22***</td>
</tr>
<tr>
<td>Satisfaction with peer relationships</td>
<td>747</td>
<td>4.77 (.78)</td>
<td>.83</td>
<td>.27***</td>
</tr>
<tr>
<td>CBCL</td>
<td>1287</td>
<td>1.68 (.44)</td>
<td>.93</td>
<td>-.23***</td>
</tr>
</tbody>
</table>

Note: STER-CY = Situational Test of Emotion Regulation Abilities for Chinese Youth; MSCEIT-YV = Mayer–Salovey-Caruso Emotional Intelligence Test – Youth Version; ERQ = Emotion Regulation Questionnaire; CDI = Children’s Depression Inventory; CBCL = Child Behavior Checklist.

---

https://doi.org/10.1017/prp.2019.31 Published online by Cambridge University Press
correlations were significant, indicating that the participants with higher emotion regulation ability exhibited lower depression and anxiety, fewer internalizing and externalizing problem behaviors, and better relationships with their friends. These results are consistent with prior research findings (e.g., McLaughlin et al., 2011; Zeman et al., 2006).

Therefore, based on the data of the current study, the STER-CY has acceptable psychometric properties and can be used to measure the emotion regulation ability of Chinese adolescents.

**Gender and grade differences**

The results showed that girls scored higher on the STER-CY than boys did, which is consistent with the prior findings that women’s emotion regulation ability was better than men’s emotion regulation ability (Rivers et al., 2012; Zhang, 2013). Some researchers found that men were more likely to use an expressive suppression strategy (Gross & John, 2003), which has been proven to be less adaptive than reappraisal (Webb et al., 2012). Moreover, women were more likely to use multiple strategies in a particular situation than men (Nolen-Hoeksema & Aldao, 2011), suggesting that women might have more strategies and be more flexible in strategy use. These findings could help us better understand why women do better in emotion regulation than men.

For grade differences, emotion regulation ability significantly increased from 5th- to 7th-grade, but it did not improve from 7th- to 11th-grade. Similar results were also found in other studies. For example, Rivers and colleagues (2012) used the MSCEIT-YV to measure EI of children aged 10–13 years and found that 11-year-old children scored relatively higher on the emotion management subscale than other age groups. Similarly, Zhang (2013) found that 7th-grade students scored higher on the emotion regulation test than 8th-, 10th-, and 11th-grade students did. The lifespan development theory emphasizes that growth and decline co-exist during the developmental process of psychological characteristics (Baltes, Lindenberger, & Staudinger, 2006), so there is a possibility that the development of emotion regulation ability might be non-linear. Stagnation and even decline could occur during the process of development, although emotion regulation ability generally increases with age because of physical maturation and life experiences (Zimmermann & Iwanski, 2014).

Another possible explanation for the grade differences is that the STER-CY could not reveal the whole picture of emotion regulation ability in terms of what it measures. The diversity and flexibility of regulation strategies are an important part of emotion regulation ability during middle childhood and adolescence (Compas, Connor-Smith, Saltzman, Thomsen, & Wadsworth, 2001). However, the static nature of SJTs has limited ability to measure them. Some information missed by the STER-CY might lead to unexpected grade differences.

**The features of the STER-CY related to Chinese culture**

All participants of this study were students in mainland China, so we cannot make cultural comparisons and examine cultural influences on adolescents’ emotion regulation. However, the situations and responses of the STER-VY could help to better understand possible cultural differences in emotion regulation that have been found by some researchers. For example, Chinese adolescents seem to prefer a downregulation of their negative emotions in public and to not let others see their grief, sadness, anger, disappointment and so on. We collected a large number of situations in which participants intended to reduce the intensity of their negative emotions in front of others (e.g., relieve grief when being punished in class; relieve disappointment when failing in an election), so this kind of situation might be relatively more frequent in the final test and might be consistent with previous findings. For instance, the ideal emotion is "calm" for Chinese individuals (Tsai et al., 2006). Chinese people are more likely to conceal their emotions rather than show their emotions to others (Butler et al., 2007). The belief of 荣辱不惊 (stay calm no matter how proud or humiliated you feel) could have a profound influence on Chinese emotion regulation goals.

In addition, some school situations might reflect the characteristics of Chinese or Asian educational environments. For example, students are strictly forbidden to speak in class; there are many exams and people care about scores and ranking; and relatively more emotional events happen in school or are related to learning because this is where students spend most of their time. These features might not be salient in the tests developed in Western culture.

**Limitation and future directions**

The development of the STER-CY would contribute to the research and practice of emotion regulation ability of Chinese adolescents, but there are some limitations in the current study. First, there are fewer extrinsic regulation items (4 items) than intrinsic regulation items (11 items). One possible reason for this is that some of the extrinsic regulation items had to be deleted because their emotion regulation goals were to manipulate others’ emotion for reasons of self-interest (e.g., enrage someone in order to win a contest) and might be against fundamental ethics principles. On the other hand, in the STER-CY, the intrinsic regulation items cover three regulation goals, while the extrinsic regulation items cover two regulation goals. This might be another reason why there are fewer extrinsic regulation items than intrinsic regulation items. Future researchers should make more efforts to generate appropriate situations of extrinsic regulation. Second, as mentioned above, SJTs could measure an individual’s ability to choose a proper strategy of emotion regulation in static situations. This index might not be enough to reflect the level of an individual’s emotion regulation ability because it cannot be seen how they think, select and change the strategy. Future studies should improve the test by adding more indices or adopt more advanced techniques to generate dynamic situations in which participants could choose and modify the strategy according to situational changes. Third, most of the data were obtained from self-reports in the current study. Behavioral data and evidence from other independent sources should be collected in future validation studies. Fourth, the recruitment of participants may limit the generalization of the current findings. Participants were all normal children who came from mainland China. Children in Hongkong and Taiwan and those with special needs should be considered in future research.

**References**


Appendix: The Situational Test of Emotion Regulation Ability for Chinese Youth

The questions below ask how to help the person in the story feel a certain way. There are four actions to choose from in the story. An effective action would be one that is useful in helping the person feel a certain way but will not bring other negative effects. Please figure out which action would be the most effective and which action would be the least effective.

1. Wangyun’s desk mate is rude. He often uses Wangyun’s pens without permission. Wangyun wants to make him feel shameful and guilty about his behavior. Which action would be the most effective to make him feel shameful and guilty? Which action would be the least effective?
   A. Tit for tat.
   B. Criticize him for his behavior.
   C. Do not lend him anything.
   D. Tell him how others feel about his behavior.

2. Xiaoyuan does not do well on the English exam. Xiaohe received the highest score in the class, so she is very proud. “I thought you are good at English,” Xiaohe said ironically. Xiaoyuan feels sad and angry about Xiaohe’s words. Which action would be the most effective to help Xiaoyuan relieve anger? Which action would be the least effective?
   A. Tell herself that it is a joke and she should not mind Xiaohe’s words.
   B. Praise Xiaohe and ask her about how to study English.
   C. Point out the areas that Xiaohe does poorly in.
   D. Talk with parents or friends about her feelings.

3. Zhenghan has planned to see a movie with his friends after school, but his teacher does not dismiss the class on time. The movie will begin soon. Zhenghan feels anxious and fidgety. Which action would be the most effective to help Zhenghan relieve anxiety and fidgetiness? Which action would be the least effective?
   A. Try not to think about the movie.
   B. Tell himself that studying is more important than seeing a movie.
   C. Remind the teacher that it is time to end the class.
   D. Think that the class will be over very soon.

4. Liyan and her parents get lost when traveling. They spend a great deal of time looking for the correct way but fail to find it. Liyan’s parents blame each other and become increasingly furious. Liyan wants her parents to calm down. Which action would be the most effective to help her parents’ anger? Which action would be the least effective?
   A. Ask them not to speak loudly in public places.
   B. Tell jokes to her parents.
   C. Tell her parents that it is not anyone’s fault.
   D. Tell them that she is hungry to distract their attention.

5. Yuliang asks his classmates not to chat in class. The teacher thinks he is the one who chats in class and punishes him. Yuliang feels wronged. Which action would be the most effective to help Yuliang relieve his grief? Which action would be the least effective?
   A. Explain to the teacher that he does not chat in class.
   B. Accept criticism and punishment.
   C. Apologize to the teacher and plan to explain later.
   D. Tell himself that the teacher’s misunderstanding is normal because he does speak in class.

6. Wangping often uses her words to hurt others. Today, she says something mean to embarrass her desk mate again. Her desk mate wants to make her feel guilty and regretful. Which action would be the most effective to make Wangping feel guilty and regretful? Which action would be the least effective?
   A. Ask Wangping to put herself in others’ shoes.
   B. Do not consider Wangping as a friend anymore.
   C. Tell Wangping that what she said makes others upset.
   D. Let Wangping know that she would be a good friend if she says something nice to others.

7. Wangli gets a high score on a difficult exam. She feels very happy and confident with herself. She wants to maintain the feeling of self-confidence. Which action would be the most effective to help Wangli maintain self-confidence? Which action would be the least effective?
   A. Tell herself that she is the best every day.
   B. Keep studying hard and give rewards to herself when making progress.
   C. Think often about this successful exam.
   D. Share her happiness with her parents and friends.

8. The teacher asks the students to complete a test before going back home. Many students finish it and go home, but Yuheng does not know how to answer the last question. He feels nervous and anxious. Which action would be the most effective to help Yuheng relieve his nervousness and anxiety? Which action would be the least effective?
   A. Take a deep breath.
   B. Tell himself that he can do it.
   C. Focus on the question without looking at others.
   D. Think that not going home is not a big deal.

9. Lily runs for a position in the class cadre. She thought that her classmates would support her, but the result is that she only received a few votes. She feels disappointed and sad. Which action would be the most effective to help Lily relieve his sadness and anxiety? Which action would be the least effective?
   A. Talk about her feelings with her parents or friends.
   B. Tell herself that she is the best.
   C. Play with her friends.
   D. Tell herself: It is not a big deal.

10. Zhaoheng’s father is unhappy because of his work. When he sees Zhaoheng watching TV, he loses his temper and scolds him, “Why do you not spend more time studying?” Zhaoheng feels sad and aggrieved. Which action would be the most effective to help him relieve his sadness and grievance? Which action would be the least effective?
   A. Stop watching TV and go study.
   B. Tell his father how he feels when his father calms down.
   C. Do something he likes, such as listening to music.
   D. Comfort his father and tell him that he has finished his homework.
11. Wangdi tells Liuyun, who is the boy she likes. Liuyun says she will not let others know, but she breaks her promise. Wangdi feels very angry and embarrassed. Which action would be the most effective to help Wangdi relieve her anger and embarrassment? Which action would be the least effective?
   A. Keep a distance from Liuyun.
   B. Tell Liuyun how she feels and ask Liuyun to explain.
   C. Think that it does not matter if it is known.
   D. Go out for a walk.

12. Liuhao is playing chess with someone. Wangbin comes to watch and says, “You are too stupid to play chess. I can do much better than you.” Liuhao wants Wangbin to feel ashamed. Which action would be the most effective to make Wangbin feel ashamed? Which action would be the least effective?
   A. Ignore Wangbin and focus on playing chess.
   B. Play chess with Wangbin.
   C. Remind Wangbin to keep silent while watching chess.
   D. Ask Wangbin to go away if he cannot shut up.

13. Wanghong has a big fight with her desk mate. Both of them are very angry. Wanghong wants to calm down and resolve the conflict. Which action would be the most effective to help her relieve anger? Which action would be the least effective?
   A. Ask the teacher or friends to help resolve the conflict.
   B. Put herself in her desk mate’s shoes.
   C. Think about the good things that have happened between her and her desk mate.
   D. Be alone for some time.

14. Wangbei is slightly absent-minded in class and gives a ridiculous answer to the teacher’s question. His classmates laugh at him. He feels awkward and ashamed. Which action would be the most effective to help him relieve his feelings of awkwardness and shame? Which action would be the least effective?
   A. Actively answer the next questions.
   B. Tell himself that this kind of thing can happen to everybody.
   C. Forget it.
   D. Be determined to concentrate in class from now on.

15. Zhangqi takes the elevator to go upstairs. The elevator suddenly stops halfway, and the telephone stops working. She is in the elevator for more than a half an hour, and nobody comes to help her. She is scared. Which action would be the most effective to help relieve her fear? Which action would be the least effective?
   A. Tell herself that there must be someone to help her.
   B. Think about the knowledge she has for an emergency situation and do what she can do.
   C. Sing.
   D. Cry.