Current Psychology
https://doi.org/10.1007/s12144-025-07660-y

The development and validation of the child self-criticism scale (CSCS)

Hajra Ashra[1] · Christopher Barnes[1] · Edward James Nairn Stupple[1] · Frances Anne Maratos[1]

Accepted: 2 March 2025

Abstract
Social and competitive environmental pressures, combined with changes in self-identity, make self-criticism especially problematic in the identity-forming phase of childhood. Yet, no robust measure of self-criticism exists for use with child populations. Here, via two studies, we present the development of the first psychometrically robust measure of self-criticism for children aged seven to 11 years: the Child Self-Criticism Scale (CSCS). In study one, an item pool was generated through: (i) a literature review of existing self-criticism measures to produce an initial pool of 104 items; (ii) eight subject matter experts' assessment of the content validity of these items; and (iii) five focus groups with 33 children in total. This resulted in a final pool of 45 items. Study 2 involved data collected at two time-points. At time-point one, 45 items were administered to a sample of 394 U.K. primary school children. Factor Analysis revealed two latent factors based on a reduced set of 15 items characterised as: Criticising self and Reassuring self. Both subscales demonstrated high internal consistency (Factor 1 α.90; Factor 2 α.82). At time point two, the 15-item CSCS was administered to 214 of the participants at time point one, alongside standardised measures of wellbeing. The CSCS showed high correlations with depression, self-compassion, self-criticism, and perfectionism, with significant correlations between CSCS subscales and validation measures. Test-retest reliability at four weeks was excellent (0.74). Overall, as a short, psychometrically robust scale, the CSCS offers a promising tool for measuring the emotional wellbeing of children and/or the effectiveness of wellbeing interventions.

Keywords Mental health · Scale development · Psychometrics · Wellbeing · Youth · Compassion

Introduction

Child and Youth mental health is in a state of crisis across the world. Globally, one in seven (14%) 10–19-year-olds experience a mental health disorder (WHO, 2021), with many theorised to stem from negative emotions towards the self (Orchard et al., 2019). Additionally, rising pressures and damaging content on social media, combined with pandemic fallout, has meant young people from generations z and alpha are increasingly struggling with *self-wellbeing* (McKinsey Health Institute, 2022). The Good Childhood Report (2023) revealed that while children in the UK were happy with their family, health and home, a significantly large proportion of children were unhappy with themselves. Indeed, three out of four (77%) children and young people were unhappy with how they look (Stem4, 2022). These findings are often attributed to high-intensity social media usage through 'upward social comparison' and unrealistic or exaggerated versions of reality. For example, Murnen et al. (2003) reported children as young as 6 years old are influenced by body image ideals presented in the media.

The prominence given to exams and academic attainment within education systems also has a negative emotional impact, with almost six out of ten (57%) young people reporting feeling stressed because of a fear of making mistakes, and 96% of young people reporting they are unable to cope with academic pressures (Department of Health, 2017). This competitive pressure, exacerbated by a focus on academic achievement, is linked with increases in stress and fear of failure, shame and exclusion (Crocker et al., 2010; Maratos et al., 2022). Therefore, challenges children face today can lead to unhealthy feelings and behaviours,

✉ Hajra Ashra
hajra.ashra@open.ac.uk

✉ Frances Anne Maratos
f.maratos@derby.ac.uk

[1] School of Psychology, College of Health, Psychology and Social Care, University of Derby, Derby, UK

Published online: 31 March 2025

including feelings of inferiority (e.g., unfavourable comparison), feelings of inadequacy, self-doubt, and criticism of the self.

Notably, self-criticism is a key antecedent of mental health disorders and negative psychosocial outcomes, and may underlie vulnerability to almost all psychopathologies (for review see Löw et al., 2020). In adults, self-criticism is defined as a self-evaluative process typically involving negative self-labelling, feelings of failure and harsh judgement of oneself (Gilbert & Procter, 2006). Whilst research on self-criticism in children is very limited, existing studies evidence self-criticism as a significant vulnerability factor for depression in childhood (e.g. Abela et al., 2007; Adams et al., 2009). For instance, a 6-month prospective study by Abela and Taylor (2003) found that self-criticism predicted an increase in depressive symptoms among children aged 8–9, and 12–13. However, to the authors' knowledge, there is no research examining the effects of child self-criticism on further psychopathologies nor general social and emotional development per se. This is potentially due to the lack of any age-appropriate measure of self-criticism for children, even though self-criticism, including negative self-beliefs and judgments, are argued to play a crucial role in the development of identity during childhood (Gilbert & Irons, 2009; Zuroff et al., 1994).

Combined with vast brain developments, childhood is key for many aspects of self and identity development, including seeking autonomy, establishing peer-group relationships, establishing a sense of belonging, developing feelings of acceptance and psychosocial wellbeing (Meeus et al., 1999). For example, Erikson (1986) has argued that throughout childhood, a child acquires different views of themselves through a variety of experiences. Thus, childhood is the time when questions around self-identity are formed, for example, "Who am I?' or "Am I good enough?" (Harter, 1982). From as young as age 5, research suggests children can establish their own identity verbally, choosing more 'good' words (e.g., good, nice) than 'bad' words (e.g., mean, bad) when referring to themselves (see Cvencek et al., 2016). Then, during middle childhood, from around age 7, children are able to focus on psychological aspects of the self (e.g., "I am a kind person" or "I don't like myself"), as opposed to physical aspects (e.g., "I have brown hair") (Harter, 1998). It is at this age, and as children enter middle childhood (7 to 11 years), that they are faced with social tasks and challenges that Markus and Nurius (1984) argue form and shape a child's self-related emotions. These include: (i) developing a relatively stable and comprehensive understanding of their self; (ii) refining their understanding of how the social world works; (iii) developing standards and expectations for one's behaviour; and (iv) developing strategies for controlling or managing one's behaviour. As a result of these tasks, by approximately seven to nine years of age, children begin using social comparisons to inform lasting self-evaluations (e.g. Harter, 2006). Importantly, as a child's awareness of self and others is refined, he or she also becomes capable of (negative) self-referential emotions. For example, from approximately age seven, children become capable of self-critical thinking because of these developing self-reasoning skills, including increased social comparisons, increased focus on self, and increased cognitive abilities (see also Gergely, 2002). Research further indicates that from middle childhood there is a temporary drop in self-esteem because of emerging self-criticism (Markus & Nurius,1984).

To expand, around middle childhood, children also become increasingly sensitive to failure which, combined with evaluation and judgment of themselves by family, peers, the education system, the media, and the child themselves, if left unchecked, can begin to lead to the child constructing a critical sense of self (e.g., Cimpian et al., 2017). If this critical sense of self becomes internalised as a persistent negative cognitive style, such as, "I hate myself" or "I feel like I am a failure", it can leave the child open to future problems of self-wellbeing and mental health (e.g., depression, social anxiety; Ashra et al., 2021a; Zaccari et al., 2024). Accordingly, Bailen et al. (2019) have further revealed that, by comparison with adults, children experience more frequent and higher-intensity negative (and positive) emotions.

Consistent with the behavioural research, functional neuroimaging studies further confirm that transitions of childhood are a key period for brain maturation in the context of self-emotions. For instance, studies have shown that during childhood explicit self-referential evaluations increase activity in the medial prefrontal cortex (mPFC) as part of a broader network (Dégeilh et al., 2015). Notably, this occurs more in children (e.g., those aged 9–10-years) compared to adults (23–31-years) (Pfeifer et al., 2009), with Ray et al. (2009) noting self-referential brain activity increases from age seven to 13.

Taken together, the above literature demonstrates the need for the development of a child-appropriate measure of self-criticism. To date, when attempting to explore self-criticism in children, the lack of any appropriate measures has led to researchers using generic clinical measures of depression (e.g., Abela & Taylor, 2003; Stolow et al., 2016) or adapting measures of adult self-criticism via slight re-wording (e.g., Barcaccia et al., 2022; Cunha & Paiva, 2012). However, neither method is appropriate for the investigation of self-criticism in children. In overview, measures that reflect self-criticism as a multi-dimensional construct according to adult theory and understandings, without either involvement of the target population (i.e., children themselves),

nor relevant experts, are simply not valid or suitable for use with child respondents (Morgado et al., 2017; Ashra et al., 2021b). For example, Baron-Cohen et al. (2010) examined how emotional lexicons develop with age. Importantly, for the emotional term 'critical', only 6.3% of children aged between seven and eight, and 33.3% of children aged nine to ten comprehended the meaning of the word. Baron-Cohen et al. (2010) concluded that these significant changes in emotional lexicon comprehension should be considered by researchers and educators when designing developmentally sensitive measures. Thus, these findings add credence to the importance of producing a child self-criticism measure that is both developmentally appropriate and reflects the nature of self-criticism in childhood. In other words, a measure that is informed by emotional understanding, language and experience of 'self-critical' emotions in childhood and thus is developmentally appropriate, rather than a scale developed for adults that is then repurposed for children.

Thus, the purpose of this research was the development and validation of a psychometrically robust measure of self-criticism for children aged seven to 11 years– The Child Self-Criticism Scale (CSCS). The scale development process was based on well-established and comprehensive scale development and evaluation guidelines, which suggest consultation with the population the scale is developed for - in this case children - as a critical stage of the scale development process (e.g., Boateng et al., 2018; Clark & Watson, 2016; DeVellis, 2003). A flow chart of the scale development procedure is provided in Fig. 1, but in overview, in study 1 a pool of items that could be used in a Child Self-Criticism Scale was generated via several methodological steps (Hinkin, 1995; Kapuscinski & Masters, 2010). This included three phases: (i) development of a deductive conceptual framework based on a literature review of existing self-criticism measures available for adults, adolescents, and children, to produce an initial pool of 104 items; then (ii) inductive Subject Matter Experts (SMEs) to assess the content validity of the initial item pool; and (iii) child focus groups to produce a final pool of 45 items. In study 2, at time point one, this pool of 45 items was administered to a large sample of primary school children alongside a valid measure of depression. Following time-one data screening, a revised 40-item list was subjected to exploratory factor analysis (EFA) to identify items that contributed to a latent variable model. The resultant 24 items were further reduced through an iterative process applying Cronbach's alphas, item-total correlations, principal axis factoring (PAF) and parallel analysis (PA). This resulted in a finalised 15-item measure. At time point two, which was four weeks later, a subsample of children from the same primary schools completed the new 15-item CSCS a second time alongside valid measures of self-compassion, depression (self-criticism subscale) and perfectionism. This was to assess concurrent validity and test-retest reliability of the newly developed scale.

Study 1

Child self-criticism scale - item generation & selection

The aim of study 1 was to produce a comprehensive and content valid item pool to be used in a Child Self-Criticism Scale via several robust methodological steps (Hinkin, 1995; Kapuscinski & Masters, 2010). This included: Step 1 a literature search of existing self-criticism measure/items available for adults, adolescents, and children to produce an initial pool of 104 items; Step 2 content validity with SMEs; and step 3, focus groups with children to produce a final pool of 45 items.

Step 1: literature search of existing self-criticism measures/items

Overview

To identify 'self-critical' items, pre-existing self-criticism and negative emotional indicators available for adult, adolescent and child populations were reviewed. In classifying an item as self-critical, two criteria had to be met. The first was that the item needed to include negative self-evaluation. This could be a negative or harsh view of oneself, negative self-labelling, feelings of failure or high expectations of the self. The second criterion was that the item must be '*self-referential'*, because self-criticism is a self-evaluative process (Becker et al., 2011; Zinck, 2008). Specifically, the item needed to be composed of a *subject* and *reference* directed towards the self. As an example, a scale with the item 'I felt angry" was not retained, while an item "I felt angry with myself" was retained (see here Ashra et al., 2021b). Additionally, as Boateng et al. (2018) recommend that items identified should be as broad and as comprehensive as possible, to produce the initial item list both clinical and non-clinical measures of self-criticism were reviewed, as well as broader emotional wellbeing measures.

Adult measures of self-criticism

A comprehensive systematic review of self-critical measures by Rose and Rimes (2018) was used to identify scales used to measure self-criticism in adult populations. Full scales included: the Self-Critical Cognition Scale (Ishiyama

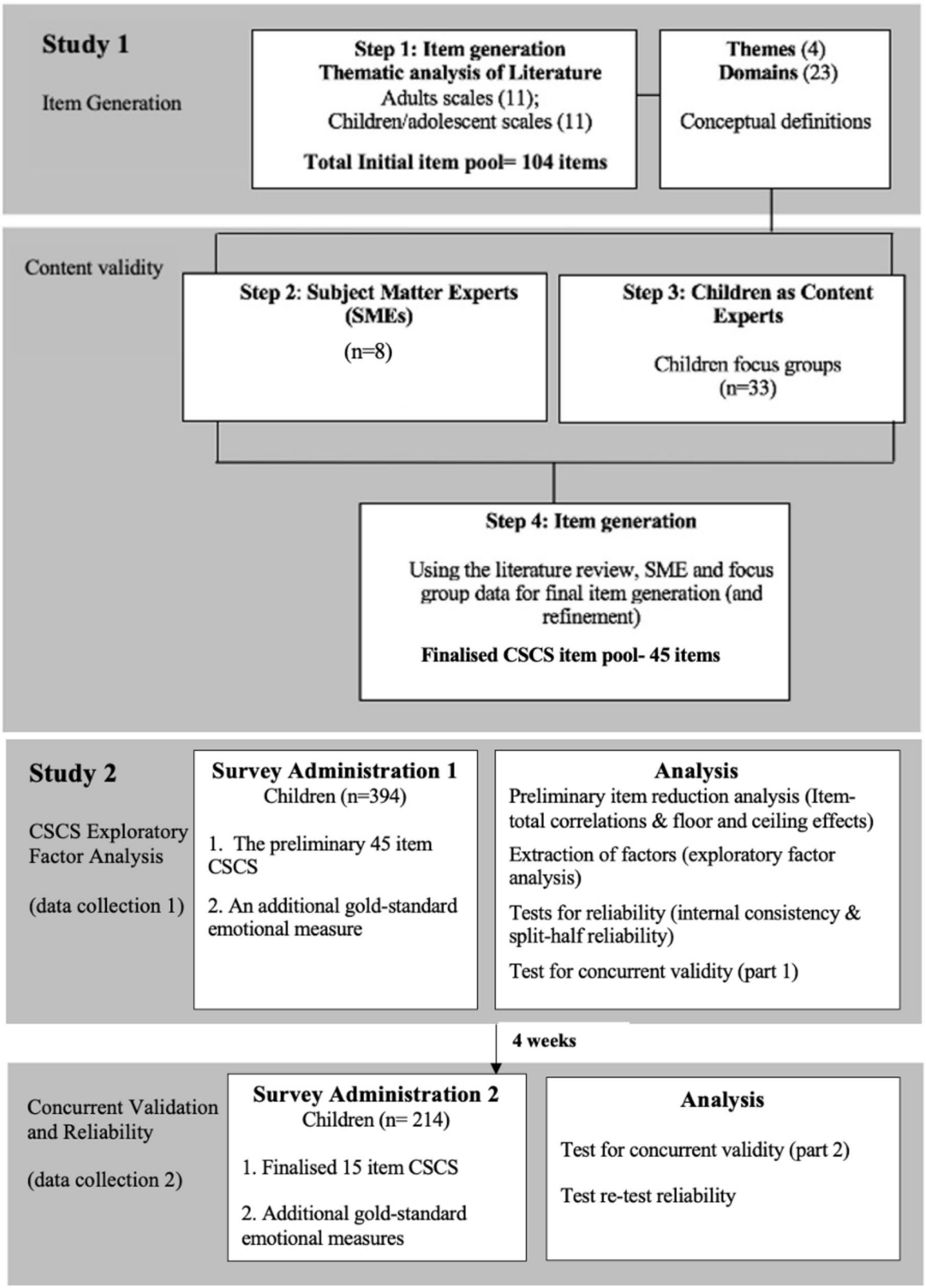

Fig. 1 The two study scale development process of the CSCS

& Munson, 1993); the Levels of Self-Criticism Scale (Thompson & Zuroff, 2004); the Forms of Self-Criticizing/Attacking and Self-Reassuring Scale (Gilbert et al., 2004); the Habit Index of Negative Thinking (Verplanken et al., 2007) and the Self-Critical Rumination Scale (Smart et al., 2016). Sub-scales included: the Attitudes Towards Self Scale (Carver & Ganellen, 1983); the Attitudes Towards Self Scale-Revised (Carver et al., 1988); the Temperament and Personality Questionnaire (Parker et al., 2006); the Self-Compassion Scale (Neff, 2003) and the Inventory of Cognitive Affect Regulation Strategies (Kamholz et al., 2006). An additional scale widely used in the literature to measure self-criticism, but not included in the review by Rose and Rimes (2018) because it is a clinical measure of depression, was also examined. This was the Depression Experiences Questionnaire (Blatt et al., 1976). This resulted in a total of 11 scales/sub-scales from inspection of the adult literature.

Children/youth measures of 'self-criticism' and emotional wellbeing

A comprehensive systematic review of self-report measures of negative self-referential emotions developed for non-clinical child and adolescent samples (Ashra et al., 2021b) was used to identify 'self-critical' items in emotional wellbeing scales already available for children and youth. All scales in the '*title and abstract screening'* stage of the systematic review were assessed for eligibility. This included 103 full-text articles reporting on 98 different measures used across clinical and non-clinical child samples. Using the self-criticism criteria overviewed above, 10 of the 98 scales included at least one 'self-critical' emotional item. These were the Depression Experiences Questionnaire-Adolescents (DEQ-A; Blatt et al., 1995); the Children's Depression Experiences Questionnaire (CDEQ; Abela & Taylor, 2003); the Children Depression Inventory-short scale (CDI; Kovacs, 1992); the Child-Adolescent Perfectionism Scale (CAPS; Flett et al., 2016); the Child & Adolescent Irrationality Scale (Bernard & Cronan, 1999); Reynolds Adolescent Depression Scale (RDAS; Reynolds, 2002); Spence Anxiety Scale-child version (Spence, 1998); the Centre for Epidemiological Studies Depression scale for children (CES- DC; Weissman et al., 1980); the Olweus Aggression Inventory for children (OAI, Ekblad & Olweus, 1986); and the Inventory of Cognitive Affect Regulation Strategies (Kamholz et al., 2006). For this scoping exercise, a wider grey literature search using 'google scholar' was also conducted. This resulted in inclusion of the Moods and Feelings Questionnaire-Long version (Angold & Costello, 2001). Thus, 11 scales from the child literature were inspected for items.

Filtering, initial item pool generation and item domains

Measures were scrutinised to select any 'self-critical' item. At this stage, those items meeting the criteria (i.e., a negative evaluation of the self, which is also self-referential in nature), were entered into an excel spreadsheet for further analysis. This resulted in a total of 183 'self-critical' items from the adult, adolescent and child scales. Thematic analysis was then used to identify underlying domains of self-criticism at the latent (vs. semantic) level (Nunnally & Bernstein, 1994; Spector, 1992).

Themes were developed iteratively and reflexively (Braun & Clarke, 2006) to fully capture domains of self-criticism within the literature. To expand, to develop themes, items were actively read, and when patterns were observed, items were grouped together to create 'domains' within a theme. As an example, the domains '*Disappointment*' and '*Anger*' were grouped together within an over-arching theme: '*Negative feelings towards the self'*. Initial grouping domains and the selection of appropriate conceptual labels was conducted by the lead author (HA). However, in accord with Morgado et al., (2017), a panel comprising the research team (FM, CB & ES) further reviewed items assigned to each domain and theme to determine the most appropriate concept labels and item fit. In addition, due to the large number of items, when similarly worded items were observed (e.g., "*I felt I was a bad person*" and "*I feel I am bad*"), one was emitted. To ensure consistency, judgments for item elimination were completed through group discussion (HA, FM, CB & ES). Reverse items of the same domain were, however, included. For example, "*I remember and dwell on my failings* and "*I don't often worry about the possibility of failure*" were both included in the initial pool of items to ensure the most relevant item/s could be chosen during the content review step. This initial item analysis process resulted in a total of 104 items identified from the 183 possible items identified.

The 104 items were grouped into four themes of negative feelings towards the self (10 domains), self-critical response to failure and self-blame (seven domains), self-oriented perfectionism (four domains) and self-reassurance (two domains). Supplementary Table 1 provides theme, domain ($n=23$) and item (n104) information.

Theme 1: negative feelings towards the self

This included a total of ten domains: self-criticism; disappointment; unworthiness; hopelessness; self-anger; self-hate; shame; self-disgust; dwelling on negative feelings towards oneself and, physical displays of negative feeling towards oneself. More broadly, these items are all concerned

with one's negative view of the self. Thus, this domain was conceptualised as "*Negative views of oneself, that are self-generated. This is a self-evaluative process involving negative or judgemental thoughts about oneself*".

Theme 2: self-critical response to failure

A total of seven domains were included as part of this theme. They were: fear of failing; inadequacies; weaknesses; making mistakes; blaming oneself for faults, failures, mistakes and weaknesses; being too hard on oneself; and inferiority. Items in this domain are concerned with an individual's tendency to blame oneself when things go wrong, which may lead to feelings of failure and inferiority. Accordingly, this theme was conceptualised as "*Critically blaming oneself for faults, mistakes and weaknesses".*

Theme 3: self-oriented perfectionism

This theme was composed of four domains: perfectionism; expectations of oneself; high standards; and, evaluations from others. Items in this theme comprised beliefs that striving for perfection and being perfect are important. Items in this domain are characterised by setting excessively high standards and having a "perfectionist motivation". This theme was conceptualised as "*A strong motivation to be perfect, including setting and striving for unrealistic standards and/or striving to avoid criticism/rejection from others".*

Theme 4: self-reassurance

This theme comprised of two domains, positive thinking and forgiving oneself. Broadly, items are characterised by an individual's ability to remain positive following setbacks, appreciate their own weaknesses without being overly critical of oneself, being able to look at the positive side of situations and being able to forgive oneself. Consequently, this theme was conceptualised as "*Positive and warm feelings towards oneself, including acceptance, compassion, understanding of shortcomings and/or failures."*

It should be noted that some conceptual overlap existed between the themes. For example, items in the theme '*self-critical response to failures and self-blame*' also include generic negative feelings towards oneself. This overlap in themes is common and found in previous scale development papers during the scale development process (e.g., Bearss et al., 2016).

Step 2: content validity with subject matter experts (SMEs)

Participants

To assess content validity, eight subject matter experts (SMEs) were recruited to (i) ensure that the conceptual framework represented the construct of interest, and (ii) enable the elimination of unsuitable items (Kapuscinski & Masters, 2010; Ladhari, 2010; Nunnally, 1967). These were 4 males and 4 females: 6 from the UK and 2 from the USA. The SMEs included mental health clinicians, experts in child mental health; experts in Compassion Focused therapy (CFT) and Compassion Mind Training (CMT), and experts in scale development. Years of professional experience ranged from 4 to 25.

Procedure

Ethical approval was obtained from the Health, Psychology and Social Care (HPSC) Ethics Committee at the University of Derby before data collection commenced. Initially, individual e-mails were sent to 12 experts. Following positive replies and informed consent from eight of these experts, the Delphi method (Dragostinov et al., 2022) was used to establish item consensus. Each expert was asked to give their professional judgment on every item within each core domain. Using a 3-point scale, they were asked to make a judgement on each of the 104 items as a measure self-criticism in children according to three criteria: (i) appropriateness, (ii) representativeness and (iii) clarity.

The appropriateness of an item reflected if the item was: 1= 'Essential', 2= 'Useful but not essential' or 3= 'Not necessary' in the context of measuring child self-criticism. Representativeness referred to the degree to which each item represented the larger domain of interest (Grant & Davis, 1997) where 1 = representative, 2 = the item needs revision to be representative or 3 = the item is not representative of the self-criticism domain specified. Item clarity was assessed by asking participants whether each item was well written, distinct and of appropriate reading level for children aged seven and over. Responses the experts could provide were: 1= 'Yes, clear for both children and adolescents', 2= 'Clear for adolescents but needs revision for children' or 3='Not clear for children nor adolescents'. The clarity judgement assessment also included information from a prior linguistic analysis, which was progressed to highlight item words that could not easily be understood by seven-year-olds (Baron-Cohen et al., 2010). In the document provided to the SME's any such words were highlighted to the SME panel, with suggestions of alternative wording and the question of whether those alternatives were appropriate.

Data analysis and results

For item retention, expert item analysis followed the guidelines of Hardesty and Bearden (2004) using a "sum-score decision rule". This is the total score for an item across all judges, with a lower sum-score representing the most appropriate item. Given there were eight SMEs, scores could range from eight to 24. The 11 items identified as most appropriate by the SMEs were selected for discussion in the child focus groups. These included items were from all four domains and included: *I am often critical of myself; I often put myself down; Making mistakes is okay because I can learn from them; I am kind to myself; I only feel like a good person if I do well; I don't worry what other people think of me; I like myself; I find it easy to forgive myself; I get mad at myself when I make a mistake; I think I am silly when I fail something important; When things go wrong I blame myself.* Where necessary, items were amended based on the advice of the SMEs to ensure they were suitable for children. For example, the item '*I think I'm a total fool when I fail at something important*" was amended to '*I think I am silly when I fail something important'*. This is because the SMEs advised '*total fool'* is not common language used by children/young people in the UK nor the USA.

Step 3: focus groups with children

Participants

Ethical approval was obtained from the HPSC Ethics Committee at the University of Derby. Five focus groups were conducted with 33 school children aged 7–11 years (17 girls and 16 boys) from five primary schools in the East Midlands, UK. Scott (2011) recommend optimal group sizes should range from five to eight participants. Each focus group consisted of four to eight children from the specific primary school. Additionally, each focus group contained at least one pupil from each primary-age year group (i.e., ages 7–8, 8–9, 9–10 and 10–11 years), and schools were selected to represent a range of socio-economic backgrounds using the Index of Multiple Deprivation (IMD). The IDM is a UK measure of deprivation, ranging from 1 (most deprived) to 10 (least deprived). The IMD scores for the primary schools in this study were: School 1: IMD = 5; School 2: IMD = 4, School 3: IMD = 2; School 4: IMD = 8 and School 5: IMD = 7. School 3 included a large proportion of students who spoke English as an additional language. Thus, the focus groups consisted of children representing a diverse range of ages, socioeconomic statuses and gender. Children were excluded from participating in a focus group if they had a currently diagnosed mental health problem and/or had special educational needs.

Table 1 Scenarios selected for the focus groups with children

Scenarios
You take part in a school play in front of the school. You forget your lines.
You did a spelling test. You had practiced, but you got 1 out of 10.
You are drawing a picture of a cat. You are not happy with your drawing.
You are playing catch and keep dropping the ball.
You are doing your maths learning. You cannot solve the first problem.
Your friends are playing games. They do not play with you.
You fell over in the playground. Your friends start to laugh at you.
You did not get invited to your friend's birthday party.
You take a sweet from the sweet jar, without asking.

Procedure

Focus groups were conducted in a quiet location in each school and lasted 40–60 min. Each focus group began with an icebreaker, after which children were presented with cue cards of various hypothetical self-threatening scenarios that tapped feelings of self-criticism, inadequacy, failure and/or self-reassurance. This phase was essential because new themes/domains can arise from focus group discussions (De Leeuw, 2011). The full list of scenarios is presented in Table 1, and an example cue card is shown in Fig. 2a. The scenarios were developed based on potential everyday situations that a child between the ages of seven and 11 might experience, using language appropriate for this age group (e.g. "*You take part in a school play in front of the school. You forget your lines".*). A practicing and leading child clinical psychologist (MW) was asked to review the scenarios and emoticons developed, with respect to language, terminology and feedback instructions/wording. The scenarios were then 'piloted' with three psychologists (FM, CB & ES) and four children aged seven to 11, with amendments made where necessary, prior to their use in the child focus groups.

In the research phase proper, children in the focus groups were asked to specify "unhelpful thoughts" and "helpful thoughts" towards themselves when discussing a scenario using the happy, kind, neutral, angry and/or sad face emoticons as an aid. These emoticons are from the 'Smilyometer', a validated toolkit to measure emotions in children (Read et al., 2002).

Following scenario discussions, the children were introduced to the 11 items identified as most appropriate by the SMEs in the content validity phase. This part of the focus group involved pre-testing the items through 'cognitive interviewing' (De Leeuw, 2011) using '*think-aloud*' procedures. Think aloud procedures enable researchers to not only understand children's thoughts about items (e.g., as a concept), but also enable researchers understanding of the

Fig. 2 Legend: (**a**) Example of a scenario cue card shown to children as part of the focus group; (**b**) The Emoticons

everyday *language* children use; in this case to express self-critical emotions. In this part of the focus group, children were asked to read the items aloud to facilitate any sources of confusion, misunderstanding and/or the detection of language and comprehension problems. Preferences for which word/phrase was most appropriate/easiest to understand were explored. Every item from the list of 11 was discussed at least once in at least one focus group.

Response styles to items in a questionnaire measure were also discussed with the children. Of note, whilst Likert scales are commonly employed in child scale measures as they are easy and quick to administer (for review see Mellor & Moore, 2014), the number of response points is important to consider in terms of validity. To expand, whilst response points typically vary from three to five, Chambers and Johnston (2002) have suggested that younger children (5–11 years) tend to endorse responses at the extreme end of scales when presented with a limited number of items. Furthermore, response scales with a broader range of options are associated with greater responsiveness, reliability, and validity (Leung, 2011). Consequently, the validity of a ten-point Likert scale anchored with: *Not true for me; Sort of true for me; Really true for me* was discussed with the children. Additionally, two response styles were investigated, one of which was a Likert scale in the form of a number line and the other a Likert scale using demarcated boxes. Children were asked to comment upon *"Which one do you like the most?"*.

Finally, due to the potentially emotive nature of the focus groups, to close each group, children were shown a fun 'cat video' suitable for this age group.

Data analysis

In keeping with Braun and Clarke (2006) all transcripts were inductively coded for meaning and patterns enabling a set of overarching themes. Focus groups were transcribed verbatim to ensure that their original meaning was maintained. Following this, transcripts were actively read, patterns within the data searched, and domains generated. This was an iterative process. Domains were grouped together, and these groups were then reviewed and revised, while considering the meaning of these groups within the overall data. These groups were then developed into themes. The first author led the thematic analysis and kept a reflective journal to ensure transparency of the research process. However, the wider research team (HA, FM, CB & ES) reviewed all themes and domains, with any modifications discussed (e.g., moving domains to a different theme). The themes and domains identified were used as the foundation to produce a final pool of items for Study 2: scale validation and psychometric evaluation.

Child focus group results

Themes and domains that emerged from the focus groups with children, including a detailed description of each of the themes, and its relevance to the literature search conducted

in step 1 are presented below. The qualitative analysis resulted in five broad themes containing 30 domains: Internalising negative evaluation from others (5 domains); Negative feelings towards the self (6 domains); Self-critical response to failure (9 domains); Self-oriented perfectionism (2 domains); and Self-reassurance (8 domains). Supplementary Table 2 provides an overview of the five self-criticism themes, sub-ordinate domains contained within each theme generated from text extracts, and how the thematic analysis, text segments from the focus groups, and SME recommendations informed item generation.

In general, children spoke about self-criticism in a fairly similar manner to each other, but differed in their rhetoric as to experience of certain aspects of self-criticism and the importance as to different aspects of self-criticism. An important theme of self-criticism prominent in all focus groups with children, but not a distinct theme in the prior literature search (phase 1) or SME feedback (phase 2), was "*internalising negative evaluation from others"*. The further four themes tapped into similar overreaching themes as emerged from the literature search and SME content validation stage. A summary of each theme including domains examples and, where relevant, how the theme differed from the literature search is included below.

Theme 1: internalising negative evaluation from others

This theme emerged entirely from the focus group phase. Children in all focus groups expressed concerns of being ridiculed if they failed, which in turn made them feel bad about themselves, whether that be through ridicule from their peers, teachers, or parents. Here, the focus was largely on an unfavourable comparison of oneself as compared to others, evidently through physical displays of criticism from other people '*making fun'* of them, '*laughing at them'* or '*thinking*' badly of them when they made or make a mistake or fail at something. Children also worried what people might *say* about them for not doing well at something, and the concern of letting others down. Children very commonly compared their failure to others, for example, Easy-E stated, "*If I saw all my friends*, *getting it right*, *I would be angry at myself*, *because I couldn't do it*" (Easy-E, maths problem scenario).

Theme 2: self-critical response to failure

This theme taps into similar domains as theme 2 in the literature search. However, there was a greater emphasis on 'self-blame' from children for not '*trying hard enough'* or '*working hard enough'*. For example, for the play scenario, Carrie blamed herself for not revising hard enough "*I would feel like gosh! I should have remembered them; I should have revised them and I should have. It's my fault*!" (Carrie: forget play lines scenario). Additional domains for this theme included 'giving up' following setbacks/failures. For example, Lucy stated "*I just want to give up and do nothing else*" for failing a recent Maths test in school despite working hard. Ruminating over mistakes and failures was also prominent in the focus groups.

Theme 3: negative feelings towards the self

In this theme, domains mapped onto similar concepts as theme 1 in the literature search. However, the language used to express these negative feelings towards the self varied in children. For example, '*self-criticism*' was expressed in a wide range of ways. Indeed, consistent with existing literature, *anger* and *disappointmen*t towards the self were common phrases used by children to express these negative self-feelings. Children also insulted themselves, for example Geoffrey stated "*I'm a clown… I'm just stupid*". However, hopelessness, self-disgust, and self-hate, all of which were domains in the literature search, were not expressed in the child focus groups. Additional domains identified in this theme included shame, frustration with the self, guilt, and embarrassment.

Theme 4: self-oriented perfectionism

This theme mapped onto similar domains as theme 3 of the literature review. However, only two domains were identified as opposed to four domains in the literature search. These included (i) perfectionistic motivation and (ii) failure in living up to personal expectations. Perfectionist motivation was characterised by unhappiness with not achieving perfectionism. For example, Carrie recalls being upset with a recent test she had in school, despite it being near perfect "*yeh*, *I got upset when I got 18 out of 20 on the test*". Failure in living up to personal expectations was reflected by Anakin "I *would feel pretty ashamed of myself*, *that I couldn't do it and that I knew I could have done it and could have done it well*".

Theme 5: self-reassurance

This theme tapped into theme 4 of the literature search but included a much larger set of domains (i.e., eight domains as opposed to two identified in the search). Children commonly spoke about being "*proud*" and "*respecting"* themselves, even in failure. In addition, making mistakes was opined as okay because you "*tried your best*", "*have another chance*" and you can "*keep on trying*". These were all common phrases used by children. Children also often spoke about "*putting things behind you*". For example, in failing

a spelling test Chris stated "*well…. it doesn't matter really! I can still succeed in the future! Don't let things behind you worry you*, *focus on the future".*

Response style

The majority of children ($n=26$; 78%) preferred the 'number line' Likert-style scale. On further probing, it became clear that this style of responding was something they were very familiar with using in primary school settings.

Item list

The focus group interviews revealed that children interpreted the majority of the 11 items as had been conceptualised. However, for one item "I *am often critical of myself*", children between the ages of seven and nine were not familiar with the word '*critica*l' (consistent with Baron-Cohen et al., 2010). Thus, alternative wordings was considered with the children. The focus group discussions revealed a range of words used to refer to 'self-criticism' (i.e., *mad*, *harsh*, *nasty*, *horrible*, *annoyed*, *thinking badly*, *not happy*, *unkind*, *hard*, *sad*, *worried*, *unconfident*, *downhearted* and *frustrated*). However, the most common and preferred phrase was '*angry*' and/or '*mean*' towards the self.

Step 4: final item generation and refinement

An item list for the CSCS was generated by combining data and results from all prior steps. Figure 3 summarises how data from each of these steps was used to generate a CSCS 45-item pool for Study 2 (from a now reduced pool of 66 items). To expand, in developing the item list for Study 2, items were modelled on the language children used during the focus groups (Haynes et al., 1995; Brod et al., 2009). Many of the items emerged from direct quotes from the child focus group participants. Additionally, where possible, items were kept short and simple to minimise cognitive load. A major challenge throughout this process was to adapt items as needed, keeping in mind the essence of the original theoretical construct underlying content, whilst ensuring appropriateness for children aged seven to 11. An example of this was the modification of the item '*I remember and dwell on my failings'* to *'I can easily move on when I make a mistake'* to suit the language used by children aged seven to 11 as well as the SMEs recommendations. The phrasing of items that included the word '*mad*' were also amended because the child focus group data revealed children to conceptualise mad as being crazy rather than the intended 'angry with self' *(e.g.*, *"I would say that mad is crazy so when someone is… they are crazy*", Mandy).

Several practical measurement issues were further considered. For example, ethical issues that items may pose when children defined self-criticism as insulting the self were removed or amended (e.g., "*you could say 'I am stupid' 'I wish I was never born' 'I wish I was dead'*, Geoffrey). These amendments included rewording the relevant items to be appropriate for children from a general (i.e., non-clinical) population, e.g. '*I call myself mean names'; 'I often put myself down*'. Finally, common phrases used to refer to '*self-criticism'* were also re-examined and the item/s most preferred by children (e.g., '*I am often mean to myself'*) selected.

The result of this stage was the exclusion of a further 21 items resulting in a finalised item pool of 45. Readability of the 45 items was checked using the Flesch-Kincaid Grade level index. This revealed all items to be of grade 3 level. So, suitable for children of age seven to eight years. The final list of items generated for validity and psychometric testing is presented in Supplementary Table 2 (final column).

Study 1 discussion

The aim of study one was to develop an item pool that could be used in a Child Self-Criticism Scale for children aged between seven and 11 years; with items sourced in accord with best practice recommendations (e.g., Boateng et al., 2018; Clark &Watson, 2016). To this end, the findings of an exhaustive literature review, the subject matter experts (SME) content validity stage, and the focus groups with children, resulted in 45 items to be included in Study 2 (CSCS Exploratory Factor Analysis, Concurrent Validity and Reliability).

Analysis of the child focus group data supported a multidimensional conceptualisation of self-criticism broadly consistent with previous literature. The 45 items covered 28 domains, which could be subsumed within five broader themes of: internalising negative evaluations from others (6 items); self-critical response to failure (13 items); negative feelings towards the self (10 items); self-oriented perfectionism (7 items); and self-reassurance (9 items). Given the comprehensive item development procedure followed, these 45 items can be argued to have good content validity, represent the experiences and understanding of the non-clinical child population, be age-appropriate (i.e., were mainly generated using the words and phrases used by the children themselves), and pose minimal respondent burden (i.e., items are generally short and simple). Importantly, attention was focused on child perspectives of self-criticism and the children as experts, with both the former key components of the final 45 item list generated to form the initial CSCS item pool in Study 2.

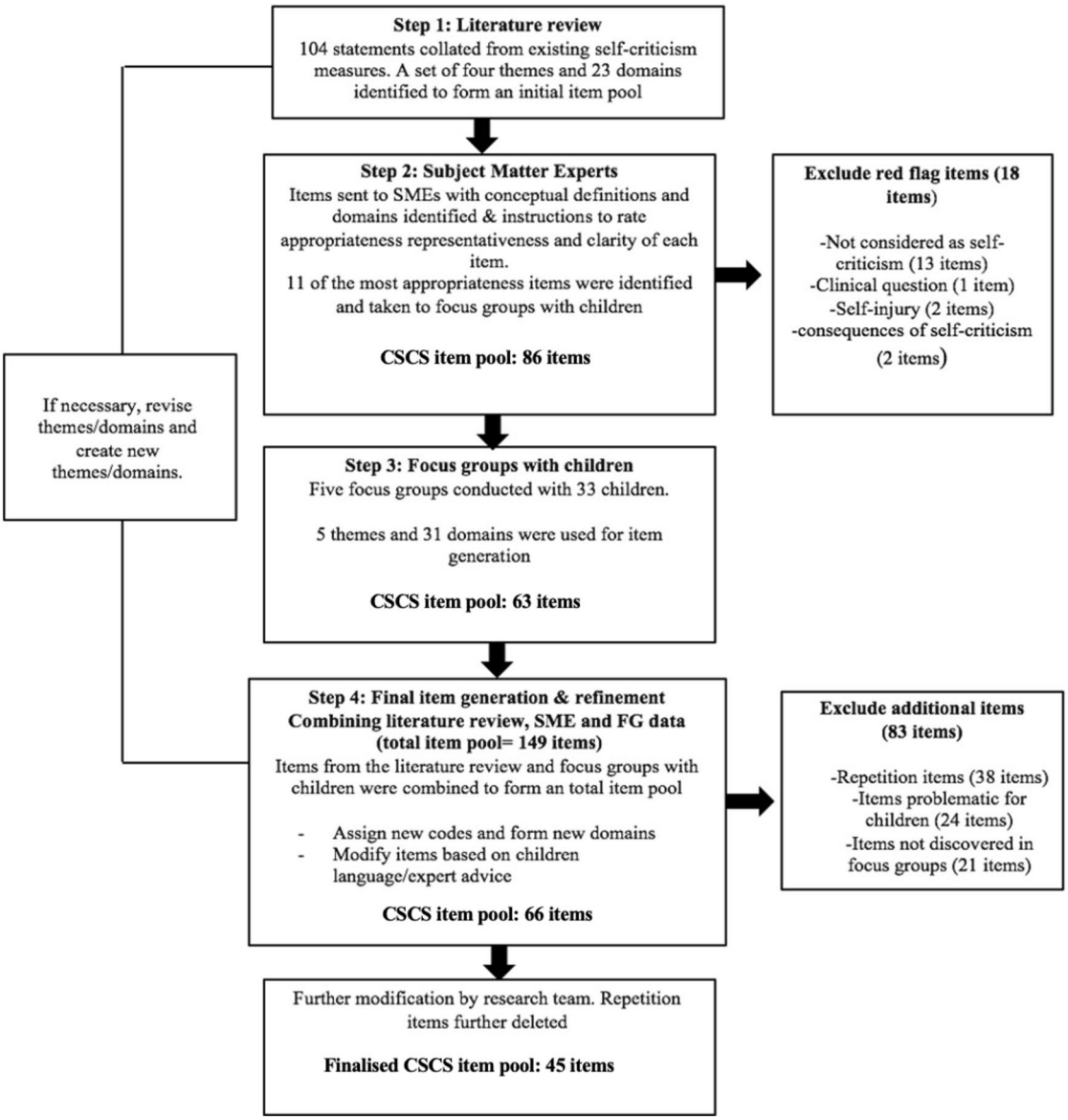

Fig. 3 Schematic flow chart demonstrating CSCS item generation

To date, taking such a rigorous scale development approach for young populations is scarce. Rather, simply modifying adult measures for younger populations using only SMEs, without the consultation of young people, is common practice (Ashra et al., 2021b). However, this study demonstrates why utilising the latter approach is poor practice. For example, if children as experts had not been consulted, the entire domain of child self-criticism as 'internalising negative evaluations from others' would have been missed. Additionally, for 'self-reassurance' being '*proud*', '*respecting self*'' and '*learning from mistakes*' were all domains identified in the child focus groups but not the original literature review. Finally, recruiting children as experts revealed that certain domains identified in the original literature review were not relevant in child conceptualisations of self-criticism. For example, '*self-disgust*' was not a concept children aged between seven to 11 recalled experiencing in their day-to-day lives, nor were they familiar with this concept. This is understandable since the development of self-criticism continues to emerge and be refined in the content of a child's beliefs about themselves (e.g., cognitions), their abilities and language. These, in turn, are informed by, and in response to, life experiences and emotional development (Cimpian et al., 2017). Thus, certain domains identified in adult literature are naturally irrelevant to children aged seven to 11.

The child focus groups also revealed how expression of certain concepts is different for children than adults. For example, for the *'self-criticism'* domain, children used a range of alternative lexicons to refer to self-critical emotions,

with the most common and preferred phrase being '*angry*' and '*mean*' towards the self. This finding is consistent with prior research demonstrating that only 6% of children aged between seven and eight, and 33% of children aged nine to ten comprehend the meaning of the word 'critical' (Baron-Cohen at al., 2010). These variations in phrasing used by children may not have been found if only the SME phase had been relied upon for item generation.

Thus, phase 1 resulted in a contextual and ecologically valid 45-item pool suitable for psychometric testing, to enable development of the first *age-appropriate* and *practical* measure of self-criticism for use with non-clinical child populations.

Study 2

CSCS exploratory factor analysis, concurrent validity and reliability

The aims of study 2 were to: (i) refine the 45-item pool developed in study 1, (ii) identify its factor structure (Time 1 analyses), and (iii) investigate reliability (Time 1 and 2 analyses) and validity of the measure (Time 2 analyses). To meet the first two aims, factor analysis was used to refine the 45-item pool to only include the most psychometrically valid and reliable items in the CSCS at Time 2. Additionally, to meet the third aim, a combination of correlational and repeated measures design testing across the two-time points was utilised. Internal consistency of all subscales was assessed against thresholds proposed by George and Mallery (2003).

Participants

Data was collected over two time points for the purposes of test-retest reliability.

Time point 1 A total of 394 children (214 girls and 180 boys) took part in the study. All children were aged between seven and 11 years (M = 9; SD = 1.11). Schools were selected to be representative of a range of socio-economic backgrounds across the East Midlands and were from a range of different IMD postcodes. The IMD scores for the primary schools in this study were: School 1: IMD = 5; School 2: IMD = 4; School 3: IMD = 2; and School 4: IMD = 9. School 3 included a large proportion of students who spoke English as an additional language.

Time point 2 A total of 214 children (115 girls and 99 boys) took part in the study, age ranged from seven to 11 years (M = 8.6; SD = 0.96). These were the same children from Time-Point 1, from School 1, School 2 and School 3.

At both time points children were excluded if they had a current diagnosed mental health problem and/or had special educational needs. The research received full ethical approval from the HPSC Ethics Committee at the University of Derby.

Materials

Child self-criticism scale (CSCS)

The newly developed CSCS was administered in paper format. At Time one a 45-item version was used, at time 2 following factor analysis, a reduced 15-item version was used. Children were asked to mark their response on a ten-point Likert scale 'number line' (1 = not true for me, 5 = sort of true for me and 10 = really true for me). All psychometric properties related to this scale are reported in the study 2 results section.

Children's depression inventory- short form (CDI: S); Kovacs, 1982)

The CDI is one of the most widely used self-report instruments of depressive symptoms for children and adolescents aged between seven and 17 years. The CDI: S consists of ten items that quantify: sadness, pessimism, self-deprecation, self-hate, crying spells, irritability, negative body image, loneliness, lack of friends, and feeling unloved. Each item is presented as a series of three phrases. Respondents are asked to select the phrase that best represents how they feel (e.g., "*I am sad once in a while*", "I *am sad many times*" or "*I am sad all the time*"). Each response is assigned a value from 0 to 2. Five items begin with the sentence that represents the greatest symptom severity and, for the further five, the sequence of choices is reversed. Scores range from 0 to 20, with a higher CDI-S score representing a higher depressive state. Previous research has demonstrated the CDI-S possesses high reliability and validity (e.g., Kovacs, 1981; Weiss et al., 1991). Its Cronbach's α reliability coefficient has been reported as 0.80 (Kovacs, 2003). The CDI: S correlates $r = .89$ with the full CDI, with concurrent validity and test–retest reliability adequate (Kovacs, 2003; Giannakopoulos et al., 2009). In the current study (Time 1), an α value of 0.86 was obtained indicating good internal consistency.

Children depression experiences questionnaire (CDEQ; Abela & Taxel, 2001)

The CDEQ was developed to assess interjective depression for children and early adolescents. The measure includes two subscales: dependency and self-criticism. Despite the measure being developed for use with clinical child populations, it is currently the only measure available that potentially measures the construct of child self-criticism. The dependency and self-criticism subscales of the CDEQ each contain five items. However, only the self-criticism sub-scale was used in this study (e.g., "*If I am not good at everything I do, I get mad at myself*"). For each statement, a child is given the following three choices: (a) not true for me, (b) sort of true for me, and (c) really true for me. Each response is assigned a value from 0 through to 2, with higher scores indicating higher levels of self-criticism. Self-criticism composite scores can range from 0 through to 10. Abela and Taylor (2003) examined the reliability and validity of the CDEQ in children aged 12 to 13 years and reported an α value of 0.61 for the self-criticism subscale, indicating questionable internal consistency. In the current study (Time Point 2), an α value of 0.60 was obtained for the self-criticism sub-scale, again indicating questionable internal consistency.

Self-compassion scale for children (SCS-C; Sutton et al., 2018)

The SCS-C is a measurement of self-compassion for children aged between eight and 12 years, adapted/re-worded from the adult Self-Compassion Scale—Short Form (SCS-SF; Raes et al., 2011). It includes 12 items addressing each of the six components of Neff's (adult) definition of self-compassion: Self-kindness, Self-judgment, Common humanity, Isolation, Mindfulness, and Over-identification. Respondents are asked to respond on a five-point Likert-type scale ranging from 1 (*Never*) to 5 (*Always*). Unlike the adult version, the child version was found to comprise of 2 factors (termed positive self-compassion and negative self-compassion). Sutton et al. (2018) found the SCS-C to have good internal consistency (Cronbach's alpha = 0.81 and–0.83). They reported the scale has strong positive associations with a general sense of self-concept ($r = .50$), optimism ($r = .45$), empathetic-related responding ($r = .42$), and weak negative associations with depression ($r = -.22$) and anxiety ($r = -.13$). In the current study (Time Point 2), an α value of 0.65 was obtained for the negative self-compassion subscale and α value of 0.81 for the positive self-compassion sub-scale, indicating questionable and good internal consistency, respectively, for the sub-scales.

Child–adolescent perfectionism scale—short form (CAPS–SF; Bento et al., 2019)

The CAPS–SF is a short version of the widely used Child–Adolescent Perfectionism Scale (CAPS) to measure perfectionism in children and adolescents. The CAPS has been reported as the most psychometrically rigorous measure of negative self-referential emotional measures for non-clinical children and adolescent populations (Ashra et al., 2021b). It consists of a four-item self-oriented perfectionism subscale (e.g., *I want to be the best at everything I do*) and a five-item socially prescribed perfectionism subscale (e.g., *My family expect me to be perfect*). Children are asked to respond on a five-point Likert-type scale, ranging from 1 (False) to 5 (True). Internal consistency of the CAPS–SF was found to be high (self-oriented perfectionism subscale α = 0.84; socially prescribed perfectionism subscale: α = 0.86; Bento et al., 2019). The test–retest correlation coefficient for the total score has been reported as 0.67 ($p < .001$). Bento et al. (2019) have further reported positive total score correlations with anxiety ($r = .212$; $p < .01$), depression ($r = .234$; $p < .01$), and stress ($r = .223$; $p < .01$). In the current study (Time Point 2), an acceptable α value of 0.72 was obtained for the socially prescribed perfectionism subscale and a questionable α value of 0.60 was obtained for the self-oriented perfectionism subscale.

Procedure

Following parental, school and child consent, all children completed paper formats of the relevant questionnaires. These were the 45-item CSCS followed by the Child Depression Scale (CDI) at Time Point 1, and at Time Point 2, a revised 15-item version of the CSCS followed by the SCS-S, the CAPS-SF and the CDEQ self-criticism subscale. Questionnaires to explore concurrent validity were split between the two time-points to minimise child fatigue and cognitive load.

At each time point, the questionnaires were completed with children in their usual classroom of circa 25–30 pupils per classroom, with the first author (HA), teacher and teaching assistant present. For Year Three and Year Four pupils (aged 7 to 9), each questionnaire item was read aloud verbally by the researcher including the response options provided. Children were instructed to complete the questionnaire independently without discussing their answers with their peers. For Year Five and Year Six pupils (aged 9 to 11), children completed the questionnaires independently, with support available when required. Questionnaires took between 20 and 45 min to complete.

In accordance with scale development guidance, a four-week interval between questionnaire administrations was utilised (Streiner & Norman, 2014).

Data analysis

For the 45-item CSCS, data were screened for missing values and outliers (e.g., for acquiescence bias). Of the 394 respondents, 25 were missing 1 or 2 item responses for a total of 33 missing data points. Missing responses represented 8% of the total data set. Given that this represented less than 10% of the data, and this is considered as small in guides for dealing with missing data, mean imputation was employed (Schumacker & Lomax, 2015). Acquiescent responding was investigated by screening for participants who responded in the same direction to both standard and reversed items. In total data from four (1.02%) of the 394 participants who took part were identified to display acquiescent responding. Their data was deleted leaving 390 data sets for further analysis.

The univariate normality of individual items and total scores was then tested. This was accomplished by investigating means, skewness, and kurtosis. This information was used to assess multivariate normality because existing tests of multivariate normality are known to be overly sensitive (Tabachnick et al., 2007). As recommended by DeVellis (2012), item-means were examined to ensure they were close to the central range. To expand, if a mean is near the range extreme, then the item may fail to detect certain values of the construct. Corrected item-total correlations were calculated for each item and those that correlated less than 0.25 with the total were removed (Everitt et al., 2006; Nunnally & Bernstein, 1994). Floor and ceiling effects were examined for each of the items, using the item means and item frequency tables. Problematic items (i.e., CSCS items with extreme means and a high percentage of participants selecting extreme scale points) were removed.

Exploratory factor analysis (EFA) was then used to identify a latent variable model for remaining questionnaire items. Both the Kaiser (1960) Eigenvalue > 1 rule and scree plot data were used to explore if there was a clear sharp decline in factor Eigenvalues (see Cattell, 1966). Furthermore, factors extracted were compared to data simulated using Horn's (1965) parallel analysis (PA) to calculate the empirical data compared to its corresponding mean simulated Eigenvalue from the random data. This process generates a correlation matrix for the dataset and extracted a principal Axis Factoring model with Eigenvalues. Henceforth, these were compared to Eigenvalues extracted from 100 iterations of random simulated data. Principal axis factoring (PAF) was used to extract factors due to detected non-normality, and direct oblimin rotation was used to aid interpretation of the model.

In order to locate items that did not strongly load onto any of the factors in the model, the 'pattern matrix' loadings were examined. Items with a low loading (i.e., items > 0.45) were removed. Items displaying "cross loading" (i.e., significant loadings onto more than one factor) were also removed. This process was continued until a simple factor structure was achieved. Finally, Cronbach's alpha and split half reliability coefficients were calculated for the overall scale and for each subscale detected through EFA.

To explore the concurrent validity of the CSCS, correlations with the self-criticism 5-item CDEQ sub-scale, the CDI:S, the CAPS-SF and the SCS-C were performed. A Pearson product-moment correlation matrix was calculated because data were normally distributed. Correlation coefficients of $\rho < 0.25$ were considered as small; 0.25–0.50 as moderate; 0.50–0.75 as good; and > 0.75 as excellent.

Temporal stability was analysed by the test–retest correlation method (Pearson correlation). A correlation of > 0.70 represents high reliability.

Finally, the readability score of the final (15-item) CSCS was explored using the Flesch readability score.

Results

Preliminary data screening

Standardised skewness and kurtosis scores were calculated as z-scores. Values for skewness and kurtosis are considered acceptable in order to prove normal univariate distribution if between −2 and +2 (George & Mallery, 2019). Several items had Z skewness and Z kurtosis score that were over these thresholds. Thus, the dataset was treated as non-parametric. Item-total correlations were conducted prior to the factor analysis. Based on recommended thresholds from Nunnally and Bernstein (1994) and Everitt et al. (2006) four items were removed due to item-total correlation coefficients below 0.25. This indicated that these items were not strongly associated with the overall construct measured as compared to other potential scale items. Subsequently, the total scores and corrected item-total correlations were then recalculated for the 41 remaining items. This revealed that all remaining items had corrected item-total correlations above 0.25, indicating they were all at least moderately correlated with a single construct. Examination of floor and ceiling effects revealed one item exhibited a low mean score and large percentage of extreme scale scores ($M = 8.14$; $SD = 3.01$). Thus, this item was deleted, resulting in 40 items for the factor analysis.

Factor analysis

SPSS Statistics version 26.0 was used to perform the EFA. Inspection of univariate descriptives for the 40 items identified substantial skew and kurtosis for some items, indicating that the data was multivariate and non-normal. Therefore, principal axis factoring (PAF) was used for factor extraction. The Kaiser Meyer Olkin measure (0.954) and Bartlett's Test of sphericity (Bartlett's chi square (990) = 8162.408, $p < .001$) indicated that the data were suitable for factor analysis. The low off-diagonal values in the anti-image correlation matrix provided further evidence that the data were suitable for factor analysis (Tabachnick et al., 2007). Parallel analysis (PA; Horn, 1965) was used to identify the number of factors to extract. This implied that two factors had Eigenvalues greater than the 95th percentile.

Two factors were also extracted from the dataset using a direct oblimin rotation. The pattern matrix revealed 14 items without factor loadings above Comrey and Lee's (1992) 'fair' criterion of 0.45. Hence, these items were removed. A further two items exhibited "cross loading" above the recommended 0.32 (Costello & Osborne, 2005) and were removed before repeating factor extraction with the remaining 24 items.

Following this iterative process, inter-item correlations were examined to identify items that were highly correlated. Where two items correlate above 0.80, Hair et al. (2006) recommend one of the two items should be removed. This process revealed four items that were highly correlated with a further four items. In each case, the correlating item with the lower factor loading was removed. For the now remaining 20 items, Cronbach's α, if the item were to be deleted, and corrected item-total correlations, were recalculated. This revealed Cronbach's α would increase if three further items from Factor 1 were removed. These items also exhibited the lowest inter-item correlations. Thus, three further items were deleted. Checks for normality were performed for the remaining 17 items. Two items showed a high level of skewness. Therefore, these two items were further deleted, resulting in a final 15 items.

In the resulting model these 15 items all loaded at > 0.45 on to one of the two factors with no cross-loadings above the recommended criteria of 0.32. Rotated factor loadings for the final model are presented in Table 2. The factors were interpreted by examining the content of items. Factor one accounted for 43.78% of the variance and consisted of nine items labelled as 'Criticising self'. Factor 2 accounted for 11.72% of the variance and consisted of six items labelled as 'Reassuring self'. For the final version of the 15 item CSCS refer to the supplementary materials (page 12–15).

Internal consistency

The items in Factor 1 and Factor 2 demonstrated excellent and good internal consistency at time point 1 respectively (Factor 1 Cronbach's α = 0.90; Factor 2 Cronbach's α = 0.82). The overall CSCS Cronbach's α was 0.90, which is also excellent. The high reliability scores suggest that the 15 items collectively, and the individual subscales, are internally consistent. Items were randomly divided into two sub-sets (part one included eight items; part two included seven items), and a Spearman Brown split-half co-efficient

Table 2 Pattern matrix of rotated factor loadings for finalised 15 item list

Item No	Item	Factor 1	Factor 2
41	I worry everyone will laugh at me when I make a mistake	.837	
36	I feel embarrassed about my mistakes	.754	
17	I feel useless when I fail something important	.730	
37	I often think that bad things only happen to me	.715	
29	I am easily disappointed with myself	.687	
33	When I don't do as well as I hoped to, I get angry with myself	.665	
32	I often feel angry with myself	.633	
8	I often put myself down	.631	
7	I am often mean to myself	.557	
40	Making mistakes is okay because I can learn from them		.818
20	It's okay if I fail as long as I tried my best		.685
45	I believe in myself		.633
43	I do not give up when I get things wrong		.600
22	I am able to put my mistakes behind me		.579
18	I find it easy to forgive myself		.512

Table 3 CSCS full scale and sub-factor scale correlations with further child emotion measure

Measure	Factors	CSCS total	CSCS Factors	
			Criticising self	Reassuring self
Depression (CDI:S) (n=394)	CDI:S total	.81**	.75**	.64**
Self-criticism subscale (CDEQ) (n=214)	Self-criticism subscale total	.57**	.56**	.40**
Self-compassion (SCS-C) (n=214)	SCS-C total	-.83**	-.76**	-.74**
	Positive Self-compassion	-.60***	-.44**	-.75**
	Negative Self-compassion	-.77**	-.79**	-.51**
Perfectionism (CAPS-SF) (n=214)	CAPS-SF total	.38**	.36**	.15**
	Self-oriented perfectionism	.58**	.37**	.24**
	Socially-prescribed perfectionism	.22**	.28**	.05

P <.01** *CSCS* Children Self-criticism scale;*CDI:S* Children's Depression Inventory- Short Form; *CDEQ* Children Depression Experiences Questionnaire; *SCS-C* Self-compassion Scale for Children; *CAPS-SF* Child– Adolescent Perfectionism Scale—Short Form

conducted. This demonstrated a significant high correlation (r=.90) further inferring that the test halves are highly correlated, and that the measure has excellent internal consistency.

Concurrent validity

The correlations between the CSCS dimensions, and each further measure (and its factors, where applicable), are shown in Table 3. The 15-item CSCS showed an excellent significant positive correlation with the CDI: S (depression). The 15-item CSCS also demonstrated a good positive correlation with the CDEQ self-criticism subscale, and a moderate significant positive correlation with the CAPS-SF (measuring perfectionism), especially the self-oriented perfectionism subscale. The CSCS further demonstrated an excellent significant negative correlation with the SCS-C (self-compassion scale). Significant correlations were not observed between the reassuring self CSCS subscale and the CAPS-SF socially prescribed subscale. Potential explanations for this finding are discussed in the discussion section.

Test re-test reliability

The test–retest correlation coefficient was excellent for the 15-item CSCS total score (r=.74, p<.001). The test–retest correlation for the criticising self CSCS subscale and reassuring self CSC subscale were both also excellent (r=.75, p<.001, r=.72, p<.001, respectively). This indicates high stability of the CSCS and its factors.

Readability

The 15-item CSCS has a Flesch readability score of Grade three for children aged seven to eight years, this translates as meaning 'Easy to read'.

Study 2 discussion

The purpose of Study 2 was the psychometric evaluation of the item set produced as a consequence of Study 1. This included item reduction techniques to derive the most reliable and valid subset of items. This process resulted in a robust scale with 15 items across two factors that demonstrated excellent psychometric properties.

The two factors were identified as Criticising Self and Reassuring Self. These were named based on an inspection of the items. To expand, items mapping onto the first factor reflected different types of self-critical thoughts and feelings including: self-anger, self-disappointment, being mean to oneself, and self-doubt. Items mapping onto the second factor reflected reassuring thoughts, feelings and believing in oneself. Items in this subscale also included reacting to mistakes and failures constructively.

Concurrent validity of the CSCS was determined with comparisons to gold-standard measures of emotional wellbeing. The CSCS demonstrated a strong significant correlation with depression. This accords with previous child literature demonstrating self-criticism is associated with depression in children aged between seven and 14 years (Blatt & Zuroff, 1992; Adams et al., 2009; Abela et al., 2007). The CSCS also correlated with perfectionism, which has not been investigated previously in this context. While the '*self-oriented perfectionism'* subscale had a good positive correlation with the total CSCS, the '*socially prescribed perfectionism*' subscale only exhibited a weak correlation. Our explanation of this is that self-oriented perfectionism is self-referential (i.e., characterised by setting excessively high standards and having a "perfectionist motivation" for oneself), whilst socially prescribed perfectionism is not (i.e., defined as the tendency for an individual to believe that others expect perfection from them) (see Enns & Cox, 2002, Hewitt & Flett, 1991). Thus, self-oriented perfectionism is mainly an

internally motivated form of perfectionism whereas socially prescribed perfectionism is mainly an *externally motivated form.* As self-criticism is a self-referential emotion (i.e., largely internally motivated), this is an expected, feasible finding that demonstrates concurrent validity of the CSCS. The CSCS also significantly negatively correlated with self-compassion. Indeed, while self-criticism is considered adverse for good emotional wellbeing, self-compassion is considered beneficial for good emotional wellbeing. Zhang et al. (2018) found a negative correlation between self-criticism and self-compassion in an adult population and Xavier et al. (2016) found such in adolescent populations. This is the first study to demonstrate this association in a child population.

To sum, study 2 demonstrates that the CSCS is a psychometrically sound measure of self-criticism in children. The concurrent validity of the measure is particularly impressive given there is currently limited research exploring the wider ramifications of self-criticism and emotional wellbeing in children, beyond its association with depression. Taken together, these positive features of the measure suggest it is likely to be valuable across many educational and community settings.

General discussion

Self-criticism is a multifaceted experience that can come into being during childhood identity development and adversely affect wellbeing. Yet no current measure investigates this construct in the context of childhood experience and conceptualisation, nor in respect to child cognitive development, emotional expression, or day-to-day experiences. In this research, we have described the development and validation of the first bespoke measure of self-criticism for children aged seven to 11 years, in accordance with best practice recommendations (e.g., Boateng et al., 2018; Clark &Watson, 2016). The result is a 15-item Child Self-Criticism Scale (CSCS) measure comprised of two factors: Criticising Self and Reassuring Self, produced through a rigorous scale development process.

Self-criticism and self-reassurance emerged as separable latent factors within the CSCS in accordance with conceptualisations of self-criticism in adult populations (e.g., the FSCRS; Gilbert et al., 2004; the SCS Neff, 2003, fMRI research by Longe et al., 2010 and the factor-analytic study by Kim et al., 2020). This indicates that as with adults, in children, self-reassurance is an independent factor to self-criticism, as opposed to simply the opposite end of a self-critical dimension. However, in child populations, how facets of these components were expressed and experienced was different to adults. Indeed, the two highest loading items of the Criticising Self factor of the CSCS included '*I worry everyone will laugh at me when I make a mistake"* and *"I feel embarrassed about my mistakes*". These items are related to fear of negative evaluations because of failing or making a mistake.

Importantly, the CSCS items were developed following inductive analysis of focus group data with children. As such, potential items did not originate exclusively from the extensive literature search of previous self-criticism measures. This fear of negative evaluation concept is not well-studied in adult nor adolescent populations. While one scale, the level of self-criticism scale (Thompson & Zuroff, 2004), does include a 'comparative self-criticism' sub-scale, this sub-scale focuses on feelings of '*inferiority*', rather than a fear of negative evaluations by others for failing or making a mistake. Yet, the importance of these items in the newly developed CSCS accords with developmental research suggesting that younger children have a higher need for approval from others. To expand, their self-worth is more contingent on social approval than for older children/adolescents (Flavell et al., 2000; Tanaka et al., 2011). These influences alongside the competitive environment young people are faced with in their day to day lives (e.g., academic failures), the rates of bullying in primary schools, and negative evaluations by peers, may exert potent negative influences on mental health and wellbeing (see Salmon et al., 1998), and child self-appraisals. For example, rejection, victimization, and other stressful peer experiences are associated with self-blame, low perceived self-competence, low self-efficacy, and diminished self-worth (Boivin & Hymel, 1997; Caldwell et al., 2004). As such, the content of these two items captures an integral part of children's experiences of self-criticism; namely 'Internalising negative evaluation from others'. Until the current research, this important domain has been entirely neglected in self-criticism measurements.

The second Factor of the CSC focused on a reassuring, encouraging and compassionate attitudes towards the self (e.g., *I believe in myself; I find it easy to forgive myself*). Items in this Factor relate to responding to failure, mistakes and setbacks with a positive stance. For example, items loading onto this factor related to accepting mistakes and learning from them, not giving up despite failures, and being able to move past mistakes/failures. Failures and making mistakes are an inevitable part of a child's day-to day experience, especially within the increasingly academic competitive nature of education (Maratos et al., 2023). In fact, children report that receiving poor grades and making mistakes with homework are among the most common distressing events in their daily lives (Hakkim & Konantambigi, 2023). Of relevance, central to the conceptualisation of 'self-reassurance' in the adult literature is the ability

to reassure the self by reminding oneself of one's positive competencies and qualities in the face of setbacks and failures (Gilbert et al., 2004).

In the adult literature an important distinction is made between individuals who self-reflect on their mistakes and problems with a self-critical lens (demonstrating poorer strategies to regulate their emotions), compared with those who focus on a more understanding and reassuring stance when they reflect upon their mistakes (and therefore demonstrate a more adaptive emotion regulation strategy) (Burwell & Shirk, 2007). These differences were reflected in the CSCS 'self-reassurance' subscale and accord well with developmental studies. For instance, Heyman et al. (1992) report that by the time children reach middle childhood there are strong individual differences in response to failure. Some children show a self-critical response, including experiencing negative moods, engaging in negative self-cognitions and giving up easily. However, other children tend to show a more 'self-reassuring' pattern of response. These children are more likely to maintain a reassuring stance and to remain actively engaged in problem solving (vs. giving up), enabling them to move on from their failures more easily. To sum, the items related to reassuring self in the face of setbacks and failures is consistent with adult literature, and provides a crucial addition to the measurement of self-criticism and resilience in child populations.

The evidence for the utility of our measure is good. However, in future studies the scale needs further testing, with a confirmatory factor analysis and an examination of the extent to which the measure can be applied beyond the initial sample, to further establish if the two-factor model identified in the EFA of the CSCS display good fit in wider population testing. It is, moreover, important to test the factor structure of the CSCS with children from a diverse set of UK and international populations to assess whether the findings associated with the scale are generalisable beyond the present sample of children. This would allow researchers, educators and practitioners to accurately identify children with unusually high scores, facilitating the use of the measure as a screening instrument in the future.

To conclude, results indicate that the CSCS is a psychometrically robust and valid measure of self-criticism which is quick and easy to administer. The measure reflects self-criticism as a multi-dimensional construct, informed by child emotional understanding, developmental stages and typical 'self-critical' experiences of children aged seven to 11 years. It also demonstrates high validity and reliability. As such, the measure could prove invaluable in supporting continued research into self-criticism (and negative self-referential emotions) in non-clinical child populations. Indeed, emerging research into child wellbeing is limited by a lack of age-appropriate rigour measures, including self-criticism (Maratos et al., 2024). This is of key importance when one considers improvements in wellbeing may be mediated by decreases in self-criticism (Zaccari et al., 2024), confirming the importance of self-criticism in childhood emotional wellbeing and adjustment. Therefore, the outcome of the present research is not only the first rigorous 'adult *and* child as experts' informed measure of self-criticism, but also a measure that can tap a construct of crucial developmental and therapeutic importance. Therefore, the CSCS should have major utility in work with children.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12144-025-07660-y.

Acknowledgements Many thanks to all the pupils who participated in the study and the school staff who organised the focus groups and data collection in primary schools. We would also like to thank those individuals who formed our subject matter expert panel during the content validity phase of the scale development, including Dr Mary Welford, Dr Anastacia Kudinova, Rory Cahill, Professor Antony Spirito and Professor Aimee Aubeeluck. We also thank the two anonymous reviewers whose feedback has helped enhance the quality of the manuscript and Professor James Elander who further served as a critical colleague.

Author contributions HA and FM incepted the study. All authors contributed to the design and development of the materials for the study. HA collected all data. HA performed data analysis and interpretations for study 1, supported by FM. HA and ES performed data analysis and interpretations for study 2. The first draft of the manuscript was written by HA and FM and reviewed by CB, with ES providing further feedback. All authors approved the final manuscript.

Funding The authors have no relevant financial or non-financial interests to disclose.

Data availability The data will be made publicly available at the point of acceptance/publishing.

Declarations

Ethical approval Ethical clearance was approved by the Health, Psychology and Social Care (HPSC) Ethics Committee at the University of Derby before data collection commenced.

Informed consent Informed consent was obtained from all individual participants involved in the study.

Research involving human participants and/or animals All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee.

Competing interests The authors have no competing interests to declare that are relevant to the content of this article.

References

Abela, J. R. Z., & Taxel, E. (2001). *Personality predispositions to depression in schoolchildren.* In Annual meeting of the Canadian Psychological Association.

Abela, J. R., & Taylor, G. (2003). Specific vulnerability to depressive mood reactions in schoolchildren: The moderating role of self-esteem. *Journal of Clinical Child and Adolescent Psychology*, *32*(3), 408–418.

Abela, J. R., Sakellaropoulo, M., & Taxel, E. (2007). Integrating two subtypes of depression: Psychodynamic theory and its relation to hopelessness depression in early adolescents. *The Journal of Early Adolescence*, *27*(3), 363–385.

Adams, P., Abela, J. R. Z., Auerbach, R., & Skitch, S. (2009). Self-Criticism, dependency, and stress reactivity: An experience sampling approach to testing Blatt and Zuroff's (1992) theory of personality predispositions to depression in High-Risk youth. *Personality and Social Psychology Bulletin*, *35*(11), 1440–1451. https://doi.org/10.1177/0146167209343811

Angold, A., & Costello, E. J. (2001). The epidemiology of depression in children and adolescents. *The Depressed Child and Adolescent*, *2*, 143–178.

Ashra, H., Barnes, C., Stupple, E. J. N., & Maratos, F. A. (2021a). Negative Self-referential emotions and mental health in youth: The importance of self-criticism. *Current Research in Psychiatry*, (4), 63–67. https://doi.org/10.46439/Psychiatry.1.020

Ashra, H., Barnes, C., Stupple, E., J, N., & Maratos, F., A (2021b). A systematic review of Self-Report measures of negative Self-Referential emotions developed for Non-Clinical child and adolescent samples. *Clinical Child and Family Psychology Review*, *24*(2), 224–243. https://doi.org/10.1007/s10567-020-00339-9

Bailen, N. H., Green, L. M., & Thompson, R. J. (2019). Understanding emotion in adolescents: A review of emotional frequency, intensity, instability, and clarity. *Emotion Review*, *11*(1), 63–73.

Barcaccia, B., Hartstone, J. M., Pallini, S., Petrocchi, N., Saliani, A. M., & Medvedev, O. N. (2022). Mindfulness, social safeness and self-reassurance as protective factors and self-criticism and revenge as risk factors for depression and anxiety symptoms in youth. *Mindfulness*, *13*(3), 674–684.

Baron-Cohen, S., Golan, O., Wheelwright, S., Granader, Y., & Hill, J. (2010). Emotion word comprehension from 4 to 16 years old: A developmental survey. *Frontiers in Evolutionary Neuroscience*, *2*(1), 109, 1–8. https://doi.org/10.3389/fnevo.2010.00109

Bearss, K., Taylor, C. A., Aman, M. G., Whittemore, R., Lecavalier, L., Miller, J., Pritchett, J., Green, B., & Scahill, L. (2016). Using qualitative methods to guide scale development for anxiety in youth with autism spectrum disorder. *Autism*, *20*(6), 663–672. https://doi.org/10.1177/1362361315601012

Becker, J. C., Tausch, N., & Wagner, U. (2011). Emotional consequences of collective action participation: Differentiating self-directed and outgroup-directed emotions. *Personality and Social Psychology Bulletin*, *37*(12), 1587–1598.

Bento, C., Pereira, A. T., Azevedo, J., Saraiva, J. M., Flett, G. L., Hewitt, P. L., & Macedo, A. (2019). Development and validation of a short form of the Child–Adolescent perfectionism scale. *Journal of Psychoeducational Assessment*, *38*, 26–36.

Bernard, M. E., & Cronan, F. (1999). The child and adolescent scale of irrationality: Validation data and mental health correlates. *Journal of Cognitive Psychotherapy*, *13*(2), 121–132.

Blatt, S. J., & Zuroff, D. C. (1992). Interpersonal relatedness and self-definition: Two prototypes for depression. *Clinical Psychology Review*, *12*(5), 527–562.

Blatt, S. J., D'Afflitti, J. P., & Quinlan, D. M. (1976). Experiences of depression in normal young adults. *Journal of Abnormal Psychology*, *85*(4), 383.

Blatt, S. J., Zohar, A. H., Quinlan, D. M., Zuroff, D. C., & Mongrain, M. (1995). Subscales within the dependency factor of the depressive experiences questionnaire. *Journal of Personality Assessment*, *64*(2), 319–339.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, *6*, 149.

Boivin, M., & Hymel, S. (1997). Peer experiences and social self-perceptions: A sequential model. *Developmental Psychology*, *33*(1), 135.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Brod, M., Tesler, L. E., & Christensen, T. L. (2009). Qualitative research and content validity: Developing best practices based on science and experience. *Quality of Life Research*, *18*(9), 1263–1278.

Burwell, R. A., & Shirk, S. R. (2007). Subtypes of rumination in adolescence: Associations between brooding, reflection, depressive symptoms, and coping. *Journal of Clinical Child and Adolescent Psychology*, *36*(1), 56–65. https://doi.org/10.1207/s15374424jccp3601_6

Caldwell, M. S., Rudolph, K. D., Troop-Gordon, W., & Kim, D. Y. (2004). Reciprocal influences among relational self-views, social disengagement, and peer stress during early adolescence. *Child Development*, *75*(4), 1140–1154.

Carver, C. S., & Ganellen, R. J. (1983). Depression and components of self-punitiveness: High standards, self-criticism, and overgeneralization. *Journal of Abnormal Psychology*, *92*(3), 330.

Carver, C. S., La Voie, L., Kuhl, J., & Ganellen, R. J. (1988). Cognitive concomitants of depression: A further examination of the roles of generalization, high standards, and self-criticism. *Journal of Social and Clinical Psychology*, *7*(4), 350–358. https://doi.org/10.1037/0021-843X.92.3.330

Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, *1*(2), 245–276.

Chambers, C. T., & Johnston, C. (2002). Developmental differences in children's use of rating scales. *Journal of Pediatric Psychology*, *27*(1), 27–36.

Cimpian, A., Hammond, M. D., Mazza, G., & Corry, G. (2017). Young children's self-concepts include representations of abstract traits and the global self. *Child Development*, *88*(6), 1786–1798.

Clark, L. A., spsampsps Watson, D. (2016). Constructing validity: Basic issues in objective scale development. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research* (pp. 187–203). American Psychological Association. https://doi.org/10.1037/14805-012

Comrey, A. L., & Lee, H. B. (1992). *A first course in factor analysis*. Erlbaum.

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment Research and Evaluation*, *10*(1), 7. https://doi.org/10.7275/jyj1-4868

Crocker, J., Canevello, A., Breines, J. G., & Flynn, H. (2010). Interpersonal goals and change in anxiety and dysphoria in first-semester college students. *Journal of Personality and Social Psychology*, *98*(6), 1009.

Cunha, M., & Paiva, M. J. (2012). Text anxiety in adolescents: The role of self-criticism and acceptance and mindfulness skills. *The Spanish Journal of Psychology*, *15*(2), 533–543.

Cvencek, D., Greenwald, A. G., & Meltzoff, A. N. (2016). Implicit measures for preschool children confirm self-esteem's role in maintaining a balanced identity. *Journal of Experimental Social Psychology*, *62*, 50–57.

De Leeuw, E. (2011). *Improving data quality: Cognitive and social development and its role in questionnaire construction and pretesting*. Retrieved from https://www.aka.fi/globalassets/awanhat/documents/tiedostot/lapset/presentations-of-the-annual-seminar-10-12-may-2011/surveying-children-and-adolescents_de-leeuw.pdf

Dégeilh, F., Guillery-Girard, B., Dayan, J., Gaubert, M., Chételat, G., Egler, P. J., & Viard, A. (2015). Neural correlates of self and its interaction with memory in healthy adolescents. *Child Development*, *86*(6), 1966–1983.

Department of Health (2017). Transforming children and young People's mental health provision: A green paper. *UK Department of Health and Education*. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/664855/Transforming_children_and_young_people_s_mental_health_provision.pdf

DeVellis, R. F. (2003). *Scale development: Theory and applications* (2nd ed.). Sage.

Devellis, R. (2012). *Scale development theory and applications*. Sage Publications.

Dragostinov, Y., Harðardóttir, D., McKenna, P., Robb, D., Nesset, B., Ahmad, M., & Rajendran, G. (2022). Applying the Delphi method as an important step in psychometric scale development.

Ekblad, S., & Olweus, D. (1986). Applicability of olweu'aggression inventory in a sample of Chinese primary school children. *Aggressive Behavior*, *12*(5), 315–325.

Enns, M. W., spsampsps Cox, B. J. (2002). The nature and assessment of perfectionism: A critical analysis. In G. L. Flett spsampsps P. L. Hewitt (Eds.), *Perfectionism: Theory, research, and treatment* (pp. 33–62). American Psychological Association. https://doi.org/10.1037/10458-002

Everitt, H. A., Little, P. S., & Smith, P. W. (2006). A randomised controlled trial of management strategies for acute infective conjunctivitis in general practice. *Bmj*, *333*(7563), 321–332.

Flavell, J. H., Green, F. L., & Flavell, E. R. (2000). Development of children's awareness of their own thoughts. *Journal of Cognition and Development*, *1*(1), 97–112.

Flett, G. L., Hewitt, P. L., Besser, A., Su, C., Vaillancourt, T., Boucher, D., & Gale, O. (2016). The Child–Adolescent perfectionism scale: Development, psychometric properties, and associations with stress, distress, and psychiatric symptoms. *Journal of Psychoeducational Assessment*, *34*(7), 634–652. https://doi.org/10.1177/0734282916651381

George, D., spsampsps Mallery, P. (2019). *IBM SPSS statistics 26 step by step: A simple guide and reference*. Routledge.

George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference. 11.0 update* (4th ed.). Allyn & Bacon.

Gergely, G. (2002). The development of Understanding self and agency. In *Blackwell handbook of childhood cognitive development*, pp 26–46.

Giannakopoulos, G., Kazantzi, M., Dimitrakaki, C., Tsiantis, J., Kolaitis, G., & Tountas, Y. (2009). Screening for children's depression symptoms in Greece: The use of The children's depression inventory in a nation-wide school-based sample. *European Child & Adolescent Psychiatry*, *18*(8), 485–492.

Gilbert, P., & Irons, C. (2009). Shame, self-criticism, and self-compassion in adolescence. *Adolescent Emotional Development and the Emergence of Depressive Disorders*, *1*, 195–214.

Gilbert, P., & Procter, S. (2006). Compassionate Mind training for people with high shame and self-criticism: Overview and pilot study of a group therapy approach. *Clinical Psychology & Psychotherapy: An International Journal of Theory & Practice*, *13*(6), 353–379.

Gilbert, P., Clarke, M., Hempel, S., Miles, N., & Irons, C. (2004). Criticizing and reassuring oneself: An exploration of forms, styles and reasons in female students. *Br J Clinical Psychology*, *43*(1), 31–50. https://doi.org/10.1348/014466504772812959

Grant, J. S., & Davis, L. L. (1997). Selection and use of content experts for instrument development. *Research in Nursing & Health*, *20*(3), 269–274.

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate data analysis*. Pearson/Prentice Hall.

Hakkim, A., & Konantambigi, R. M. (2023). Experiencing childhood in the context of societal academic expectations. *Psychological Studies*, *68*(2), 149–116. https://doi.org/10.1007/s12646-022-00656-9

Hardesty, D. M., & Bearden, W. O. (2004). The use of expert judges in scale development: Implications for improving face validity of measures of unobservable constructs. *Journal of Business Research*, *57*(2), 98–107.

Harter, S. (1982). The perceived competence scale for children. *Child Development*, *53*(1), 87–97. https://doi.org/10.2307/1129640

Harter, S. (1998). The Development of Self-Representation. In W. Damon & N. Eisenberg (Eds.), *Handbook of child psychology, social, emotional, and personality development* (Vol. 5, pp. 18–553). Wiley.

Harter, S. (2006). Developmental and individual difference perspectives on self-esteem. In D. K. Mroczek & T. D. Little (Eds.), *Handbook of personality development* (pp. 311–334). Lawrence Erlbaum Associates.

Haynes, S. N., Richard, D., & Kubany, E. S. (1995). Content validity in psychological assessment: A functional approach to concepts and methods. *Psychological Assessment*, *7*(3), 238–242.

Hewitt, P. L., & Flett, G. L. (1991). Perfectionism in the self and social contexts: Conceptualization, assessment, and association with psychopathology. *Journal of Personality and Social Psychology*, *60*(3), 456–465.

Heyman, G. D., Dweck, C. S., & Cain, K. M. (1992). Young children's vulnerability to self-blame and helplessness: Relationship to beliefs about goodness. *Child Development*, *63*(2), 401–415.

Hinkin, T. R. (1995). A review of scale development practices in the study of organizations. *Journal of Management*, *21*(5), 967–988.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185.

Ishiyama, F. I., & Munson, P. A. (1993). Development and validation of a self-critical cognition scale. *Psychological Reports*, *72*(1), 147–154.

Kaiser, H. F. (1960). The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, *20*, 141–151.

Kamholz, B. W., Hayes, A. M., Carver, C. S., Gulliver, S. B., & Perlman, C. A. (2006). Identification and evaluation of cognitive affect-regulation strategies: Development of a self-report measure. *Cognitive Therapy and Research*, *30*(2), 227–262. https://doi.org/10.1007/s10608-006-9013-1

Kapuscinski, A. N., & Masters, K. S. (2010). The current status of measures of spirituality: A critical review of scale development. *Psychology of Religion and Spirituality*, *2*(4), 191–201.

Kim, J. J., Kent, K. M., Cunnington, R., Gilbert, P., & Kirby, J. N. (2020). Attachment styles modulate neural markers of threat and imagery when engaging in self-criticism. *Scientific Reports*, *10*(1), 1–10.

Kovacs, M. (1981). Rating scales to assess depression in school-aged children. *Acta Paedopsychiatrica*, *46*(1), 305–315.

Kovacs, M. (1982). *The Children's Depression Inventory: A self-rated depression scale for school-aged youngsters*. University of Pittsburgh.

Kovacs, M. (1992). *The Children's Depression Inventory*. Manual North Tanawanda.Multi-Health Systems.

Kovacs, M. (2003). *Children's Depression Inventory: Technical manual update*. Multi-Health Systems.

Ladhari, R. (2010). Developing e-service quality scales: A literature review. *Journal of Retailing and Consumer Services*, *17*, 464–477.

Leung, S. O. (2011). A comparison of psychometric properties and normality in 4-, 5-, 6-, and 11-point likert scales. *Journal of Social Service Research*, *37*(4), 412–421.

Longe, O., Maratos, F. A., Gilbert, P., Evans, G., Volker, F., Rockliff, H., & Rippon, G. (2010). Having a word with yourself: Neural correlates of self-criticism and self-reassurance. *Neuroimage*, *49*(2), 1849–1856.

Löw, C. A., Schauenburg, H., & Dinger, U. (2020). Self-criticism and psychotherapy outcome: A systematic review and meta-analysis. *Clinical Psychology Review*, *75*., Article 101808. https://doi.org/10.1016/j.cpr.2019.101808

Maratos, F. A., Hurst, J., Harvey, C., & Gilbert, P. (2022). *Embedding compassion in schools. Applied positive school psychology*. Routledge.

Maratos, F., Byrd, J., & Mosey, C. (2023). Schooling in England–An overview. *DYNAMIS Rivista Di Filosofia E Pratiche Educative*, *5*(5), 21–33.

Maratos, F. A., Wood, W., Cahill, R., Tronco Hernández, Y. A., & Gilbert, P. (2024). A Mixed-Methods study of compassionate Mind training for pupils (CMT-Pupils) as a School-Based wellbeing intervention. *Mindfulness*, *15*, 4590478.

Markus, H. J., & Nurius, P. S. (1984). *Self-understanding and self-regulation in middle childhood. Development during middle childhood: The years from six to twelve*. National Academies Press.

McKinsey Health Institute. (2022). Retrieved from https://www.mckinsey.com/mhi/our-insights/heat-waves-the-war-in-ukraine-and-stigma-gen-zs-perspectives-on-mental-health

Meeus, W., Iedema, J., Helsen, M., & Vollebergh, W. (1999). Patterns of adolescent identity development: Review of literature and longitudinal analysis. *Developmental Review*, *19*(4), 419–461.

Mellor, D., & Moore, K. A. (2014). The use of likert scales with children. *Journal of Pediatric Psychology*, *39*(3), 369–379.

Morgado, F. F., Meireles, J. F., Neves, C. M., Amaral, A. C, & Ferreira, M. E. (2017). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão E Crítica*, *30*(0), 3.

Murnen, S. K., Smolak, L., Mills, J. A., & Good, L. (2003). Thin, sexy women and strong, muscular men: Grade-school children's responses to objectified images of women and men. *Sex Roles*, *49*, 427–437.

Neff, K. D. (2003). The development and validation of a scale to measure self-compassion. *Self and Identity*, *2*(3), 223–250. https://doi.org/10.1080/15298860390209035

Nunnally, J. C. (1967). *Psychometric Theory*. McGraw-Hill.

Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.

Orchard, F., Pass, L., & Reynolds, S. (2019). I am worthless and kind'; the specificity of positive and negative self-evaluation in adolescent depression. *The British Journal of Clinical Psychology*, *58*(3), 260–273. https://doi.org/10.1111/bjc.12215

Parker, G., Manicavasagar, V., Crawford, J. O., Tully, L., & Gladstone, G. (2006). Assessing personality traits associated with depression: The utility of a tiered model. *Psychological Medicine*, *36*(8), 1131–1139.

Pfeifer, J. H., Masten, C. L., Borofsky, L. A., Dapretto, M., Fuligni, A. J., & Lieberman, M. D. (2009). Neural correlates of direct and reflected self-appraisals in adolescents and adults: When social perspective-taking informs self-perception. *Child Development*, *80*(4), 1016–1038.

Raes, F., Pommier, E., Neff, K. D., & Van Gucht, D. (2011). Construction and factorial validation of a short form of the self-compassion scale. *Clinical Psychology & Psychotherapy*, *18*(3), 250–255.

Ray, R. D., Shelton, A. L., Hollon, N. G., Michel, B. D., Frankel, C. B., Gross, J. J., & Gabriel, J. D. (2009). Cognitive and neural development of individuated self-representation in children. *Child Development*, *80*(1), 1232–1242. https://doi.org/10.1111/j.1467-8624.2009.01327.x

Read, J. C., MacFarlane, S., & Casey, C. (2002). Endurability, engagement and expectations: Measuring children's fun. *Interaction design and children*, *2*(1), 1–23.

Reynolds, W. M. (2002). *Reynolds adolescent depression scale– second edition: Professional manual*. Psychological Assessment Resources.

Rose, A. V., & Rimes, K. A. (2018). Self-criticism self-report measures: Systematic review. *Psychology Psychotherapy Theory Research Practice*, *91*, 450–489. https://doi.org/10.1111/papt.12171

Salmon, G., James, A., & Smith, D. M. (1998). Bullying in schools: Self-reported anxiety, depression, and self-esteem in secondary school children. *BMj*, *317*(7163), 924–925.

Schumacker, R. E., & Lomax, R. G. (2015). *A beginner's guide to structural equation modeling* (4th ed.). Routledge.

Scott, A. (2011). Focussing in on focus groups: Effective participative tools or cheap fixes for land use policy? *Land Use Policy*, *28*(4), 684–694.

Smart, L., Peters, J., & Baer, R. (2016). Development and validation of a measure of Self-Critical rumination. *Assessment*, *23*(3), 321–332. https://doi.org/10.1177/1073191115573300

Spector, P. E. (1992). *Summated rating scale construction: An introduction*. Sage.

Spence, S. H. (1998). A measure of anxiety symptoms among children. *Behaviour Research and Therapy*, *36*(5), 545–566.

Stem4. (2002). Body Image among young people. Retrieved from: https://stem4.org.uk/wp-content/uploads/2022/12/Body-image-among-young-people-Negative-perceptions-and-damaging-content-on-social-media…new-survey-reveals-Dec-22.pdf

Stolow, D., Zuroff, D. C., Young, J. F., Karlin, R. A., & Abela, J. R. (2016). A prospective examination of self-compassion as a predictor of depressive symptoms in children and adolescents. *Journal of Social and Clinical Psychology*, *35*(1), 1–20.

Streiner, D. L., & Norman, G. R. (2014). *Health Measurement Scales: A Practical Guide to their Development and Use*. Oxford University Press.

Sutton, E., Schonert-Reichl, K. A., Wu, A. D., & Lawlor, M. S. (2018). Evaluating the reliability and validity of the self-compassion scale short form adapted for children ages 8–12. *Child Indicators Research*, *11*(4), 1217–1236.

Tabachnick, B. G., Fidell, L. S., & Ullman, J. B. (2007). *Using multivariate statistics.* Pearson.

Tanaka, M., Wekerle, C., Schmuck, M. L., Paglia-Boak, A., & MAP Research Team. (2011). The linkages among childhood maltreatment, adolescent mental health, and self-compassion in child welfare adolescents. *Child Abuse & Neglect*, *35*(10), 887–898.

The Good Childhood Report. (2023). Retrieved from: https://www.childrenssociety.org.uk/sites/default/files/202309/The%20Good%20Childhood%20Report%202023.pdf

Thompson, R., & Zuroff, D. C. (2004). The levels of self-criticism scale: comparative self-criticism and internalized self-criticism. *Personality and Individual Differences*, *36*(2), 419–430.

Verplanken, B., Friborg, O., Wang, C., Trafimow, D., & Woolf, K. (2007). Mental habits: Metacognitive reflection on negative self-thinking. *Journal of Personality and Social Psychology*, *92*(3), 526–541.

Weiss, B., Weisz, J. R., Politano, M., Carey, M., Nelson, W. M., & Finch, A. J. (1991). Developmental differences in the factor structure of the children's depression inventory. *Psychological Assessment: A Journal of Consulting and Clinical Psychology*, *3*(1), 38–42.

Weissman, M. M., Orvaschel, H., & Padian, N. (1980). Children's symptom and social functioning: Self-report scales. *Journal of Nervous and Mental Disorders*, *168*(12), 736–740.

World Health Organization (WHO) (2021). *Key messages*. Retrieved from https://www.who.int/key-messages

Xavier, A., Pinto-Gouveia, J., Cunha, M., & Carvalho, S. (2016). Self-criticism and depressive symptoms mediate the relationship between emotional experiences with family and peers and self-injury in adolescence. *The Journal of Psychology*, *150*(8), 1046–1061.

Zaccari, V., Mancini, F., & Rogier, G. (2024). State of the Art of the literature on definitions of self-criticism: A meta-review. *Frontiers in Psychiatry*, *15*(1), 1–18. https://doi.org/10.3389/fpsyt.2024.1239696

Zhang, H., Carr, E. R., Garcia-Williams, A. G., Siegelman, A. E., Berke, D., Niles-Carnes, L. V., & Kaslow, N. J. (2018). Shame and depressive symptoms: Self-compassion and contingent self-worth as mediators? *Journal of Clinical Psychology in Medical Settings*, *25*(4), 408–419.

Zinck, A. (2008). Self-referential emotions. *Consciousness and Cognition*, *17*, 496–505. https://doi.org/10.1016/j.concog.2008.03.014

Zuroff, D. C., Koestner, R., & Powers, T. A. (1994). Self-criticism at age 12: A longitudinal study of adjustment. *Cognitive Therapy and Research*, *18*(4), 367–385. https://doi.org/10.1007/BF02357511