Outcome measure	Difficulties assessed	Age range	Reporter	Time period in focus	Number of items	Sub-scales	How to obtain
Revised Child Anxiety Depression Scale (RCADS) [13]	Anxiety and depression	8 to 18	Young person and parent/ carer	None	47	Six subscales: - Separation anxiety disorder - Social phobia - Generalized anxiety disorder - Panic disorder - Obsessive compulsive disorder - Low mood Two total scales: - Total Anxiety - Total Internalising	All versions are available for use and can be downloaded in accordance with the terms of use at https://www.childfirst.ucla.edu/resources/. Any adaptations are not authorised without written permission from the developers.

Generalized Anxiety Disorder Assessment – 7 (GAD-7) [14]	Anxiety	13 and above	Young person	Past two weeks	7	Generalized Anxiety	Available at: www.phqscreeners.com/select-screener No permission is required to reproduce, translate, display or distribute the questionnaire.

Patient Health Questionnaire - 9 (PHQ-9) [15]	Depression	13 and above	Young person	Past two weeks	9	One sub-scale measuring general depressive symptoms, or two-subscales segregating cognitive/affective and somatic depressive symptoms.	Available at: www.phqscreeners.com/select-screener No permission is required to reproduce, translate, display or distribute the questionnaire.

Youth Self Report [16]	Emotional and Behavioural difficulties	11 to 18	Young person	Past six month	112	Six DSM-oriented scales: - Affective problems - Anxiety problems - Somatic problems - Attention-deficit-hyperactivity problems - Oppositional problems - Conduct problems Two broad band subscales: - Internalising - Externalising Under the two broad band subscales, there are eight narrow band syndrome subscales: - Rule-breaking behaviour - Aggressive behaviour - Withdrawn - Somatic complaints - Anxiety and depression - Social problems - Through problems	The Youth Self Report is under copyright. Information about use and ordering is available at https://aseba.org/.

Eating Disorders Examination Questionnaire (EDE-Q) [17]	Eating disorders	14 and above	Young person	Past four weeks	28	Four sub-scales: - Restraint - Eating concern - Weight concern - Shape concern One Total Scale	The EDE-Q is under copyright. For non-commercial research, this measure is freely available at www.credo-oxford.com/7.2.html and no permission needs to be sought. For commercial use of the EDE-Q contact: credo@medsci.ox.ac.uk

Strengths and Difficulties Questionnaire (SDQ) [18]	Emotional and Behavioural difficulties	2-18	Young person, parent/ carer, teacher	Past 6 months	25	Five subscales: - Emotional symptoms - Conduct problems - Hyperactivity/inattention - Peer relationships problem - Prosocial behaviour One Total Difficulties score	The SDQ is under copyright. Paper versions may be downloaded from Youth*in* Mind (https://www.sdqinfo.org/) and photocopied without charge by individuals or non-profit organisations, provided they are not making any charge to families. Users are not permitted to create or distribute electronic versions for any purpose without prior authorisation from Youth*in* Mind.

Outcome measure	Scoring	Scoring interpretation	Comparison data available
Revised Child Anxiety Depression Scale (RCADS)	Responses are provided on a 4-point scale ranging from 'Never' to 'Always'. The RCADS can be scored using spreadsheets or syntax developer or manually following instructions on the scoring aids, both available at www.childfirst.ucla.edu/resources/. The young person's equivalent US School Grade must be entered, which is grade one below the UK school year. A "t-score" (age- and gender- adjusted score) is calculated from the raw score (total score of the scale or subscale) to allow for comparison with normative data.	Larger scores indicate higher anxiety and depression difficulties. An overall t-score of 65 means that the score is roughly in the top 7% of scores of community samples of young people of the same age (described as borderline clinical by the developer) and a score of 70 means that the score is roughly in the top 2% of scores of community samples of young people of the same age (described as the clinical threshold by the developer). The RCADS was developed to assess symptoms of anxiety and depression in line with DSM-IV criteria. A t-score of 65 to 69 may suggest that a young person may be experiencing clinical levels of anxiety or depression (e.g., requiring support from specialist mental health services). A t-score of 70 or above may suggest the young person is likely to be experiencing clinical levels of anxiety or depression. Decisions about care or referral to specialist mental health services should be taken based on discussion with the young person and where relevant their family, clinical expertise and consultation with colleagues (ideally from multi-disciplinary perspectives), best available research evidence, and information about relevant local and national services.	Clinical thresholds available for: - Australia [23] - Denmark [24] - United States of America [22]

Generalized Anxiety Disorder Assessment – 7 (GAD-7)	Responses are provided on a 4-point scale ranging from 'Not at all' to 'Nearly every day'. Responses are then recoded as follows: 'Not at all' = 0 'Several Days' = 1 'More than half the days' = 2 'Nearly every day' =3. The scores on the seven questions are summed to give a total score between 0 and 21.	Higher scores indicate higher levels of anxiety. Total scores can be interpreted as follow: - 5 to 9: mild anxiety - 10 to 14: moderate anxiety - 15 and above: severe anxiety When GAD-7 is used as a screening tool, further evaluation is recommended when a total score of 10 or above is obtained. For example, if a young person has a score of 10 or above, this may suggest that further discussion, consultation, and/or assessment is needed to determine if support from specialist mental health services is required. Decisions about care or referral to specialist mental health services should be taken based on discussion with the young person and where relevant their family, clinical expertise and consultation with colleagues (ideally from multi-disciplinary perspectives), best available research evidence, and information about relevant local and national services.	Clinical thresholds available for Germany [27]

Patient Health Questionnaire - 9 (PHQ-9)	Answers are given on a 4-point scale ranging from "Not at all" to "Nearly every day". Responses are then recoded as follows: 'Not at all' = 0 'Several days' = 1 'More than half the days' = 2 'Nearly every day' = 3. The total score can be obtained by summing responses to the 9 items (note the additional difficulty item is not used in the calculation). Total scores can range from 0 to 27.	Total scores can be interpreted as follow for adolescent (which are slightly different than for adults): - 5 to 10: mild depression - 11 to 14: moderate depression - 15 to 19 above: moderate to severe depression - 20 or above: severe depression Higher scores on the PHQ-9 indicate higher depressive symptoms. According to Kroenke and colleagues [36], major depressive disorder "should be considered in patients who endorse ≥5 of the 9 symptoms as present "more than half the days" (the 9th item counts if endorsed "several days") and one of the first two symptoms (depressed mood or loss of interest) is endorsed" (p346). Respondents scoring on the item 9 "Thoughts that you would be better off dead or of hurting yourself in some way" must be followed up by a clinical interview to determine suicide risk. For example, if a young person has a score of 11 or above, this may suggest that further discussion, consultation, and/or assessment is needed to determine if support from specialist mental health services is required. Decisions about care or referral to specialist mental health services should be taken based on discussion with the young person and where relevant their family, clinical expertise and consultation with colleagues (ideally from multi-disciplinary perspective), best available research evidence, and information about relevant local and national services.	Clinical thresholds data are available for Germany [32].

Youth Self Report	Responses are provided on a 3-point scale ranging from 'Absent' (0) to 'Occurs often' (2). The Youth Self Report can be scored by converting subscale scores into t-scores. Please see the user manual and https://aseba.org/ for t-score cut offs for the subscales.	Higher t-scores indicate higher difficulties. Please see the user manual for instruction for scoring and interpretation. We are not able to provide further information on scoring and interpretation as it is not publicly available. Decisions about care or referral to specialist mental health services should be taken based on discussion with the young person and where relevant their family, clinical expertise and consultation with colleagues (ideally from multi-disciplinary perspective), best available research evidence, and information about relevant local and national services.	Clinical thresholds from the following countries are available at [38]: - Australia - China - Israel - Jamaica - the Netherlands - Turkey - The United States

Eating Disorders Examination Questionnaire (EDE-Q)	Responses range from "No days" to "Every day" and are then recoded as follows: 'No days' = 0 '1-5 days' = 1 '6-12 days' =2 '13-15 days' = 3 '16-22 days' =4 '23-27 days' = 5 'Every day' = 6 Four subscales and one global score can be calculated as follows: • Restraint = (Item 1 + Item 2 + Item 3 + Item 4 + Item 5) / 5; • Eating Concern = (Item 6 + Item 7 + Item 9 + Item 15 + Item 34) / 5; • Weight Concern = (Item 11 + Item 14 + Item 29 + Item 31 + Item 32) / 5; • Shape Concern = (Item 10 + Item 11 + Item 12 + Item 13 + Item 30 + Item 33 + Item 35 + Item 36) / 8; • Global Score = (Restraint + Eating Concern + Weight Concern + Shape Concern) / 4	Higher scores on the global scale and subscales denote more problematic eating behaviours and attitudes. Scores of 4 or above can be interpreted as being more likely to be in the clinical range. Decisions about care or referral to specialist mental health services should be taken based on discussion with the young person and where relevant their family, clinical expertise and consultation with colleagues (ideally from multi-disciplinary perspective), best available research evidence, and information about relevant local and national services.	Community norms have been published for female adolescents [45] and female and male young adults in the UK [46] as well as in Australia [47], in the US [48] [49] and elsewhere.

Strengths and Difficulties Questionnaire (SDQ)	Scores on the 25 items need to be recoded as follows: 'Not True' = 0; 'Somewhat True' = 1 'Certainly True' = 2. Note that items 7, 11, 14, 21 and 25 need to be revered. Scores on individual subscales are obtained by summing all relevant items; each scale ranges from 0 to 10. A Total Difficulties score can also be generated by summing all subscales apart from the prosocial subscale. Please see www.sdqinfo.org/py/sdqinfo/c0.py for more detailed information regarding the scoring procedure.	Higher numbers indicate higher difficulties (except for the prosocial scale, where higher numbers indicate higher prosocial behaviour). Instructions for interpreting the SDQ when scored by hand are available at https://www.sdqinfo.org/py/sdqinfo/c0.py. Different versions of the SDQ have slightly different thresholds. For example, using the newer 4-band categories of the self-report version for 4-17 year olds, Total Difficulties score can be interpreted as follows: - 0-14 close to average - 15 to 17 slightly raised - 18 to 19 high - 20 to 40 very high For example, if a young person has a Total Difficulties score of 15 or above, this may suggest that further discussion, consultation, and/or assessment is needed to determine if support from specialist mental health services is required. Decisions about care or referral to specialist mental health services should be taken based on discussion with the young person and where relevant their family, clinical expertise and consultation with colleagues (ideally from multi-disciplinary criteria), best available research evidence, and information about relevant local and national services.	Normative data for the following populations are available at https://sdqinfo.org/g0.html: - Australia - Denmark - Finland - Italy - Germany - Japan - Spain - Sweden - United Kingdom - United States of America

Outcome measure	Reliability	Validity	Overview of the measure	Other versions
Revised Child Anxiety Depression Scale (RCADS)	The RCADS subscales showed good internal consistency and good one-week test-retest reliability in community [13] [52] [23] [53] and clinical samples [22].	A number of studies have reported good concurrent validity between subscales of the RCADS and other measures of depression and anxiety (e.g. [13] [22]). The 6-factor structure has been supported in multiple studies, providing evidence for the structural validity of the RCADS [52] [23] [53]. There is moderate evidence of measurement invariance across different ethnic groups, in particular in the US between African American and Caucasian young people [52] and across countries [54, 55], however further research is required to explore invariance across a wider range of racial and ethnic groups.	The measure provides a reliable and valid tool to assess a range of anxiety disorders in children and young people. The evidence currently available indicate that the tool has good psychometric properties.	A 25-item version is also available for both young people and parents [56]. The RCADS is currently available in more than 15 languages.

Generalized Anxiety Disorder Assessment – 7 (GAD-7)	Good test-retest reliability between self-administered and practitioner-administered questionnaires was reported by Spitzer and colleagues [14]. Good scale reliability for outpatients with anxiety and mood disorders was found by Rutter and colleagues [57].	The GAD-7 shows strong criterion validity for identifying possible cases of GAD [14]. There is also evidence of good construct validity; Löwe et al. [27] substantiated the 1-dimensional structure of the GAD-7 and its factorial invariance for gender and age. The GAD-7 also yielded significant intercorrelations with the PHQ-2 and the Rosenberg Self-Esteem Scale in their research, suggesting good concurrent validity. However, there is evidence suggesting the GAD-7 shows poor measurement invariance across different racial groups. For example, Parkerson et al. [58], found that Black respondents tended to have GAD-7 scores that did not accurately reflect the severity of their generalised anxiety disorder.	This short questionnaire provides a reliable and valid screening tool to detect the presence and severity of generalised anxiety disorders.	Some versions of GAD-7 also include an additional question to estimate how much an individual's symptoms have affected their daily functioning. The GAD-7 is available in more than 40 languages.

Patient Health Questionnaire - 9 (PHQ-9)	Evidence indicates excellent internal consistency of the PHQ-9 in adolescent and adult populations and excellent test-retest reliability [33, 36].	Extensive evidence demonstrates the ability of the PHQ-9 to detect depression in adolescents e.g. [1] [59]. Research currently supports a one-factor structure measuring general depressive symptoms, and a two-factor structure, segregating cognitive/affective and somatic depressive symptoms [60]. Good convergent validity has also been found, with significant correlations obtained between the PHQ-9 and measures of functional impairment, emotional difficulties, self-harm and suicidality [31] [32]. Measurement invariance for the PHQ-9 was observed in the US across racial/ethnic groups in the adult [61] and adolescent population [62] and in the Netherlands [63] [64]. A recent study by Teymoori and colleagues [65] also found measurement invariance across different eight linguistic groups across European countries.	This short questionnaire is used extensively to detect the presence and severity of depressive symptoms. Overall evidence indicates good psychometric properties.	Different versions of the measure exist that include fewer items (e.g. PHQ-2) and a specific version for adolescents is also available (PHQ-A) [66]. The PHQ-9 is available in more than 50 languages.

Youth Self Report	There is evidence of good internal consistency for the broad band subscales in younger and older young people, with evidence of good internal consistency for the narrow band subscales in older youths [16].	There is evidence of good structural validity. There is evidence that young people with a diagnosis of anxiety had significantly higher scores on the anxiety DSM-oriented subscale of the YSR than young people without a diagnosis of anxiety; similar evidence was found for the DSM-oriented subscales attention problems and oppositional problems [16].	This measure, and the other versions for different respondents, are widely used. It provides a comprehensive assessment of emotional and behavioural difficulties. It is one of the longer measures we have included, but it can be broken down in a number of different ways.	Different versions of the same family of measures are available for parents/carers and teachers. The YSR is available in over 100 languages.

Eating Disorders Examination Questionnaire (EDE-Q)	Test-retest reliability and internal consistency are generally reported as good for the EDE-Q subscales in undergraduate and adult samples (for a review, see [67]).	Studies exploring the convergent validity of the EDE-Q have demonstrated good agreement between the measure and the interview-based EDE [67]. The EDE-Q global score was found to be highly accurate in discriminating individuals with an eating disorder from those without, and moderately accurate in discriminating individuals with binge eating disorder from those with obesity [68]. Other studies have also suggested that the EDE-Q could accurately discriminate between individuals with an eating disorder and those without [69]. However, there is considerable disagreement regarding the factor structure of the measure. Some studies supporting a four-factor model in adults, but not replicating the original subscales suggested by Fairburn and colleagues [67]. An increasingly number of studies have reported that a 3-factor model based on 7 items may be a better fit, for example in young adults [70] [71]. Limited evidence is currently available regarding the measurement invariance of the EDE-Q across different cultures and ethnic groups, with McEntee and colleagues [72] reporting that the 7-item 3-factor demonstrated invariance across ethnicity and gender, but not the 22-item four-factor version.	This measure provides a valid and reliable assessment of eating disorders in young people. Although the current evidence indicates good psychometric properties overall, care should be taken when using individual sub-scales as studies have found mixed results regarding the adequate factor structure of the measure.	The EDE-Q has been translated in more than ten languages. A version for adolescents (EDE-A) is also available and includes 36 items, focusing on eating disorders over the past two weeks. Other adaptations include 7-item [71] and 12-item [73] questionnaires to use in routine and session-by-session outcome assessment, and a Youth EDE-Q to use with children [74].

Strengths and Difficulties Questionnaire (SDQ)	Evidence currently available indicates that the SDQ is a reliable measure of emotional and behavioural difficulties. Good internal consistency was found by [18] [75] [76] and test-retest reliability was found to be moderate [18] [75]. However, evidence suggests low inter-rater reliability when comparing scores obtained from teachers, parents and young people [18] [77], indicating that taking into account more than one perspective can provide a more rounded picture.	The SDQ shows good concurrent validity with measures of mental health difficulties e.g. [75]. However, evidence for the validity of the subscales is limited [1] [5], with controversy regarding the validity of the 5-factor structure (e.g. [78] [79]). In addition, studies suggest poor measurement invariance across countries or across ethnic groups within the same country (e.g. [80] [81]).	The measure is a helpful screening tool to identify general emotional and behavioural difficulties in children and young people. It is used extensively in the UK and globally and shows overall good psychometric properties.	Different versions exist depending on the age of the child or adolescent assessed, and depending on the respondent (young person, parent/ carer, or teacher). In addition to the main assessment that includes 25 items, users can also complete an impact supplement comprising five additional questions which assesses the impact of the difficulties on the child's life using. A follow-up version is also available to measure the child's difficulties over the past month and includes 32 items. The SDQ is available in more than 80 languages.