Supplementary material 1: Additional information about the measures used in this study

Swedish translations of all measures were used in the current study. Unless otherwise stated, there are no specific validation studies in Swedish.

Measures of acceptability of I-BA

Treatment credibility – adolescent and parent versions

This measure includes four qualitative questions about treatment credibility, asking how well the treatment suits adolescents with depression, how much they believe this treatment will help them, to what extent they would recommend this treatment to a friend with depression, and how much they expect to improve from the treatment. Each item is scored on a 5-point Likert scale from 1–5, total range 4–20 with higher values indicating higher credibility.

Client Satisfaction Questionnaire (CSQ) – adolescent and parent version

CSQ[1] measures various aspects of satisfaction with treatment, e.g., perception of quality of treatment, if the treatment adequately addressed their needs and overall satisfaction. CSQ has eight self-rated items on a 4-point scale from 1–4, total range 8–32 with higher values indicating greater satisfaction). The scale has high internal consistency and correlates with therapists' estimates of client satisfaction[1].

Negative Effects Questionnaire-20 (NEQ-20) – adolescent and parent version

NEQ-20 is a condensed version of the original 32 item self-report questionnaire[2] for monitoring and reporting treatment related adverse and unwanted events such as not having confidence in one's treatment or that unpleasant memories have resurfaced. The questionnaire uses a 5-point Likert-scale ranging from 0 ("not at all) to 4 ("extremely") and includes an open question at the end about other possible negative or adverse events. Total range is 0–80 with higher values indicating more reported adverse events. In a psychometric evaluation, the Swedish version of NEQ-32 was found to have good internal consistency[2]. In another Swedish study, NEQ-20 did not demonstrate any bias in terms of responders' sociodemographic background and showed comparable validity for a condensed scale of 20 instead of 32 items[3].

Working alliance inventory, 6 items (WAI-6) – *adolescent and parent version*

WAI measures a participant's perceived working alliances with her/his therapist. WAI-6 was developed from the original 36-item WAI[4] and is rated on a 7-point Likert-scale from 1–7, total range 6–42 with higher values indicating stronger working alliances. In self-guided ICBT, the word "therapist" was changed to "programme".

Measures of clinical outcomes

Children's Depression Rating Scale, Revised (CDRS-R)

CDRS-R[5] is the most widely-used rating scale in clinical trials for assessing severity of depression and change in depressive symptoms with children and adolescents[5-7]. CDRS-R is a semi-structured interview-based measure modelled on the adult Hamilton Rating Scale for Depression. Item values range from 1–5 or 1–7, total range 17–113 with higher scores indicating more clinically significant difficulties. A raw score of ≥ 40 is indicative of depression, while a score of ≤ 28 is often used to define remission (minimal or no symptoms)[7]. CDRS-R has shown good internal consistency and good construct validity and is also considered a good measure of symptom change[7].

Children's Global Assessment Scale (CGAS)

CGAS[8] is a single-item scale from 1–100 that integrates psychological, social, and academic functioning in children as a measure of global functioning. The questionnaire is assessor-rated and has established validity and reliability[9].

Clinical Global Impression Scale – Severity (CGI-S)

CGI-S[10] is a single-item clinician rating of symptom severity for a specific disorder. Ratings are made on a seven-point scale range from 1 ("no symptoms") to 7 ("extreme symptoms"). CGI correlates well with established outcomes scales such as Hamilton Rating Scale for Depression and Brief Psychiatric Rating Scale[11].

Clinical Global Impression Scale – Improvement (CGI-I)

CGI-I[10] provides a clinician-rated opinion of global improvement. The measure consists of a single item about the level of improvement compared to state at admission, which is rated on a seven-point scale (1=very much improved, 2=much improved, 3=minimally improved, 4=no change, 5=minimally worse, 6=much worse, 7=very much worse). The questionnaire

has established validity and reliability[11]. Treatment response is commonly defined as a score of 1 (very much improved) or 2 (much improved)[12].

Need for further treatment – adolescent and parent version

This non-validated single-item questionnaire was created by David Mataix-Cols' research team, and it asks whether the participant considers her/himself in need of further treatment for her/his depression. The item is scored on a scale from 0 (no need for further treatment) to 4 (extensive need for further treatment).

Short Mood and Feelings Questionnaire (SMFQ) – *adolescent and parent version*

SMFQ[13] is a 13-item self-reported measure of depressive symptoms. Each item is scored on a 3-point scale (0 = not true, 1 = sometimes, 2 = true), total range 0–26 with higher values indicating more depressive symptoms. The total score is derived by summing together the values for each 13 items. The questionnaire has established validity and reliability[13 14]. According to a Swedish study, SMFQ is, with gender-based cut-offs, efficient as a screening tool in clinical adolescent populations, but not in children[15].

Work and Social Adjustment Scale (WSAS*) – adolescent and parent version*

WSAS is a 5-item child-rated scale of impaired functioning in school, everyday life, friends and social life, recreation, hobbies, family and close relationships and was adapted from the Work and Social Adjustment Scale[16 17]. Each item is scored on a 9-point Likert scale of 0–8, total score 0–40 with higher scores indicating greater impairment. In an evaluation of the Swedish translation of this scale, WSAS showed excellent internal consistency, adequate test-retest reliability and good convergent and divergent validity. WSAS is highly sensitive to change after treatment[17].

Revised Children's Anxiety and Depression Scale – Short Version (RCADS-S) – *adolescent and parent version*

RCADS-S[18] is a shortened version of the Spence Child Anxiety Scale, which is an adolescent and parent self-report measure of anxiety- and depression-related psychopathology. Only the anxiety subscales were administered, since depression is measured thoroughly by other measures. After eliminating the depression subscale, RCADS-S-C consists of 15 items, reflecting a single "broad anxiety" dimension. The four-graded scale

ranges from 0 = "Never" to 3 = "Always", total range 0–45 with higher scores indicating more anxiety symptoms. The 15-item Anxiety Total scale in the shortened version of RCADS has shown significant correspondence with anxiety diagnostic groups based on structured clinical interviews[18].

KIDSCREEN-10 Index – adolescent and parent version

The KIDSCREEN-10 Index[19] was developed from the longer KIDSCREEN-52 and is considered a valid measure for assessing an adolescent's general health-related quality of life. KIDSCREEN-10 consists of 10 items, each with a 5-level response category (1–5) and an additional question about general health. Total range is 10–50 with higher values indicating better health-related quality of life.

Insomnia Severity Index (ISI)

ISI*[20]* is brief screening measure of insomnia on a seven-item scale, each item scored 0–4, total range 0–28 points with higher values indicating more sleep disturbances. The scale is reliable and sensitive to change[20].

Affective Reactivity Index (ARI)

ARI[21] is measure of irritability that consists of six items on a scale of three (0–2) and one item on impairment due to irritability, total range 0–12 points with higher values indicating more irritability. ARI has been demonstrated to have excellent internal consistency and differentiated cases from controls in a clinic a community sample[21].

Behavioral Activation of Depression Scale – short form (BADS-SF)

BADS-SF is a 9-item self-report measure designed to track changes in proposed mediators of BA (activation and avoidance)[22]. Each item is scored from 0 (not at all) to 6 (completely), total score 0–54 with higher values indicating higher degree of activation and lower degree of avoidance. BADS-SF has two subscales, activation (focused, goal-directed activation and completion of scheduled activities) and avoidance/rumination (avoidance of negative aversive states and engaging in rumination rather than active problem solving). BADS-SF has acceptable internal consistency reliability, construct and predictive validity[22].

References

1. Larsen DL, Attkisson CC, Hargreaves WA, et al. Assessment of client/patient satisfaction: Development of a general scale. *Eval Program Plann* 1979;2(3):197-207. doi: 10.1016/0149-7189(79)90094-6
2. Alexander R, Anders K, Johanna B, et al. Negative Effects of Psychological Treatments: An Exploratory Factor Analysis of the Negative Effects Questionnaire for Monitoring and Reporting Adverse and Unwanted Events. *PLoS One* 2016;11(6):e0157503. doi: 10.1371/journal.pone.0157503
3. Rozental A, Kottorp A, Forsström D, et al. The Negative Effects Questionnaire: psychometric properties of an instrument for assessing negative effects in psychological treatments. *Behav Cogn Psychother* 2019;47(5):559-72. doi: 10.1017/S1352465819000018 [published Online First: 2019/03/15]
4. Horvath AO, Greenberg LS. Development and validation of the Working Alliance Inventory. *J Couns Psychol* 1989;36:223-33. doi: 10.1037/0022-0167.36.2.223
5. Poznanski E, Mokros, H. Children's depression rating scale-revised (CDRS-R). Los Angeles, CA: Western Psychological Services 2001.
6. Poznanski EO, Cook SC, Carroll BJ. A depression rating scale for children. *Pediatrics* 1979;64(4):442.
7. Mayes TL, Bernstein IH, Haley CL, et al. Psychometric Properties of the Children's Depression Rating Scale–Revised in Adolescents. *J Child Adolesc Psychopharmacol* 2010;20(6):513-16. doi: 10.1089/cap.2010.0063
8. Shaffer D, Gould MS, Brasic J, et al. A Children's Global Assessment Scale (CGAS). *Archives of General Psychiatry* 1983;40(11):1228-31. doi: 10.1001/archpsyc.1983.01790100074010
9. Green B, Shirk S, Hanze D, et al. The Children's Global Assessment Scale in clinical practice: an empirical evaluation. *Journal of the American Academy of Child and Adolescent Psychiatry* 1994;33(8):1158. doi: 10.1097/00004583-199410000-00011
10. Guy W. ECDEU assessment manual for psychopharmacology: US Department of Health, and Welfare 1976:534-537.
11. Busner J, Targum SD. The clinical global impressions scale: applying a research tool in clinical practice. *Psychiatry (Edgmont (Pa : Township))* 2007;4(7):28.
12. McCauley E, Gudmundsen G, Schloredt K, et al. The Adolescent Behavioral Activation Program: Adapting Behavioral Activation as a Treatment for Depression in Adolescence. *J Clin Child Adolesc Psychol* 2016;45(3):291-304. doi: 10.1080/15374416.2014.979933 [published Online First: 2015/01/21]
13. Angold A, Costello, E. J., Messer, S. C., & Pickles, A. . The development of a short questionnaire for use in epidemiological studies of depression in children and adolescents. *Int J Methods Psychiatr Res* 1995; 5:237 - 49.
14. Daviss WB, Birmaher B, Melhem NA, et al. Criterion Validity of the Mood and Feelings Questionnaire for Depressive Episodes in Clinic and Non-Clinic Subjects. *Journal of Child Psychology and Psychiatry* 2006;47(9):927-34. doi: 10.1111/j.1469-7610.2006.01646.x
15. Jarbin H, Ivarsson T, Andersson M, et al. Screening efficiency of the Mood and Feelings Questionnaire (MFQ) and Short Mood and Feelings Questionnaire (SMFQ) in Swedish help seeking outpatients. *PLoS One* 2020;15(3):e0230623-e23. doi: 10.1371/journal.pone.0230623

16. Mundt JC, Marks IM, Shear MK, et al. The Work and Social Adjustment Scale: a simple measure of impairment in functioning. *The British journal of psychiatry : the journal of mental science* 2002;180:461.
17. Jassi A, Lenhard F, Krebs G, et al. The Work and Social Adjustment Scale, Youth and Parent Versions: Psychometric Evaluation of a Brief Measure of Functional Impairment in Young People. *Child Psychiatry Hum Dev* 2020;51(3):453-60. doi: 10.1007/s10578-020-00956-z
18. Ebesutani C, Reise SP, Chorpita BF, et al. The revised child anxiety and depression scale-short version: scale reduction via exploratory bifactor modeling of the broad anxiety factor. *Psychol Assess* 2012;24(4):833-45. doi: 10.1037/a0027283 [published Online First: 2012/02/15]
19. Ravens-Sieberer U, Erhart M, Rajmil L, et al. Reliability, construct and criterion validity of the KIDSCREEN-10 score: a short measure for children and adolescents' well-being and health-related quality of life. *Qual Life Res* 2010;19(10):1487-500. doi: 10.1007/s11136-010-9706-5
20. Bastien CH, Vallières A, Morin CM. Validation of the Insomnia Severity Index as an outcome measure for insomnia research. *Sleep Med* 2001;2(4):297-307. doi: 10.1016/S1389-9457(00)00065-4
21. Stringaris A, Goodman R, Ferdinando S, et al. The Affective Reactivity Index: A Concise Irritability Scale for Clinical and Research Settings. *Journal of Child Psychology and Psychiatry* 2012;53(11):1109-17. doi: 10.1111/j.1469-7610.2012.02561.x
22. Manos RC, Kanter JW, Luo W. The Behavioral Activation for Depression Scale-Short Form: Development and Validation. *Behav Ther* 2011;42(4):726-39. doi: 10.1016/j.beth.2011.04.004