PSILOGOS

Research Article

open-access

Outcome-Selective Persistence after Group Cognitive–Positive Psychotherapy for Maladaptive Perfectionism: A Longitudinal Aggregate Analysis

Larry J. Young

ABSTRACT

Maladaptive perfectionism can sustain depressive and anxiety symptoms even when achievement standards themselves remain unchanged. Group cognitive behavioural therapy combined with positive psychotherapy (CBT-P) has therefore been used to weaken mistake-focused appraisal, doubt about action, avoidance, and self-punishment while retaining potentially adaptive striving. The present study asked a more discriminating question than whether CBT-P produced an average benefit: did the intervention generate a selective and durable response in the maladaptive and affective domains while leaving personal standards and parental-evaluation dimensions comparatively stable? Arm-level means, standard deviations, test statistics, and complete-case counts were analysed for 56 Chinese university students allocated to CBT-P ($n = 27$) or a 16-week waitlist ($n = 29$). Outcomes were recorded at pretreatment, immediately after eight weekly sessions, and eight weeks later. The analysis combined small-sample-corrected standardized separation, difference-in-differences on each scale, a retention coefficient, an endpoint-specificity ratio, and threshold perturbation tests. At post-treatment, standardized intervention advantages were 0.95 for concern over mistakes, 0.75 for doubts about actions, 0.89 for maladaptive perfectionism, 0.73 for depressive symptoms, and 0.76 for anxiety symptoms. Corresponding follow-up advantages were 0.62, 0.58, 0.66, 0.79, and 0.60. Differential improvement retained at follow-up ranged from 74.8% to 88.2%, with a mean of 79.1%. Mean separation across responsive endpoints was 4.22 times that of personal standards, parental expectations, and parental criticism after treatment and 4.48 times at follow-up. All five responsive outcomes exceeded a standardized advantage of 0.50 at both assessments, whereas none of the three comparatively stable dimensions did so. The results answer the research question affirmatively: CBT-P showed a domain-specific response that persisted for eight weeks, with the strongest maintenance in depressive symptoms and partial attenuation in perfectionism-related cognition. The analysis supports a treatment interpretation centred on changing punitive error processing rather than suppressing ambition.

Keywords: maladaptive perfectionism; cognitive behavioural group therapy; positive psychotherapy; university mental health; treatment persistence; outcome specificity

Emory National Biomedical Research Center, Emory University, Atlanta, GA, USA
lyoun03@emory.edu

– **Published:** 15 December 2024

1. INTRODUCTION

Perfectionism is not a unitary preference for high-quality work. Contemporary accounts distinguish achievement-oriented strivings from evaluative concerns characterised by fear of mistakes, uncertainty about performance, perceived criticism, and conditional self-worth. This distinction matters clinically because high standards can coexist with competence and satisfaction, whereas perfectionistic concerns show much more consistent associations with depression, anxiety, eating pathology, obsessive–compulsive symptoms, and reduced psychological adjustment. The Frost Multidimensional Perfectionism Scale (FMPS) made this multidimensional structure measurable by separating concern over mistakes, doubts about actions, personal standards, parental expectations, parental criticism, and organisation.[2] Its Chinese adaptation demonstrated that the same broad dimensions can be assessed in Chinese-speaking populations while retaining useful reliability and construct validity.[3] The resulting literature has repeatedly shown that the harmful component of perfectionism lies less in aspiration alone than in rigid self-evaluation when performance falls short.

This differentiation also resolves an apparent contradiction in the perfectionism literature. Perfectionistic strivings have sometimes been associated with persistence and achievement, yet perfectionistic concerns are associated with distress and impaired functioning. Stoeber and Otto's synthesis showed that positive and negative consequences become clearer when these dimensions are analysed separately.[4] Bieling and colleagues similarly found that maladaptive evaluative concerns, rather than standards in isolation, accounted for much of the relation between perfectionism and negative affect.[5] Later quantitative evidence strengthened that conclusion: a broad meta-analysis linked perfectionistic concerns to multiple forms of psychopathology, with particularly consistent associations for depressive and anxiety symptoms.[6] A separate meta-analysis found that perfectionism was related to depression, anxiety, and eating-disorder symptoms across clinical and nonclinical samples.[7] Burnout research adds a functional consequence, because concern-laden perfectionism is associated with exhaustion even where striving is not uniformly harmful.[8] These findings imply that an effective intervention need not lower every perfectionism score. It may be more clinically coherent to reduce punitive error appraisal and action doubt while allowing personally meaningful standards to remain.

University students are a particularly relevant population. Academic environments combine frequent evaluation, uncertain transitions, competitive comparison, and identity development. Students who equate mistakes with personal failure may respond to these pressures through overchecking, procrastination, avoidance, reassurance seeking, or cycles of overwork and collapse. Cluster research has shown that maladaptive perfectionists report poorer self-esteem and lower satisfaction with academic performance than adaptive perfectionists and nonperfectionists.[12] Perfectionism can therefore impair functioning even when grades or objective achievement appear satisfactory. The burden is not limited to academic productivity: repeated self-criticism may amplify low mood, physiological arousal, and uncertainty about decisions, thereby narrowing the student's ability to recover from ordinary setbacks.

Cultural context further shapes the meaning of performance standards. Research with Chinese university students in Taiwan found that perfectionism was related to both psychological well-being and achievement motivation, but the pattern varied by dimension.[10] Cross-cultural analysis has cautioned against treating parental expectations or interdependent values as intrinsically pathological; their effects depend on whether expectations are experienced as support, obligation, criticism, or a condition of belonging.[11] Work with Chinese schoolchildren has likewise shown that adaptive and maladaptive forms can be distinguished rather than reduced to a single high-score continuum.[9] For intervention design, this means that simply teaching students to "lower their standards" may be culturally insensitive and therapeutically imprecise. A more defensible target is the rigid rule that mistakes negate personal worth, coupled with the behavioural avoidance that prevents corrective learning.

Cognitive behavioural therapy (CBT) addresses this target through functional analysis, identification of automatic thoughts, behavioural experiments, cognitive restructuring, exposure to imperfection, reduction of safety behaviours, and relapse planning. The clinical model treats perfectionism as self-maintaining: self-worth becomes overly dependent on meeting demanding standards; success is discounted or followed by a higher standard; perceived failure activates self-criticism; and avoidance or overcompensation prevents a balanced appraisal. A review of adaptive and maladaptive perfectionism concluded that these interacting cognitive and behavioural processes provide a strong rationale for direct treatment.[13] Early single-case work demonstrated that cognitive behavioural methods could reduce perfectionism and associated distress,[20] while group case-series evidence suggested that psychoeducation and CBT could be delivered efficiently to several participants at once.[21]

Controlled studies subsequently established that perfectionism is modifiable. A group CBT trial reported reductions in clinical perfectionism and related symptoms.[15] Brief cognitive behavioural work with students showed that a compact intervention could produce meaningful improvement in maladaptive perfectionism,[18] and an adolescent programme demonstrated that change could be tracked longitudinally.[19] Guided internet-delivered CBT also reduced perfectionism, indicating that some treatment principles can survive changes in delivery format.[16] Dose and prescription were examined in another

internet intervention, which found benefit across delivery variants while also showing that more support can affect engagement.[22] Acceptance and commitment therapy has provided an alternative route by changing the relationship to perfectionistic thoughts rather than disputing every thought's content.[17] Across approaches, systematic review evidence indicates that psychological interventions can reduce perfectionism as well as associated depression and anxiety.[14] The common implication is that perfectionism is not an immutable personality feature, although particular dimensions may change at different rates.

Positive psychotherapy offers a complementary contribution. Conventional CBT is well suited to detecting distorted appraisals and testing rigid predictions, but severely self-critical participants may interpret corrective exercises as another performance demand. Positive psychotherapy can broaden attention toward strengths, hope, gratitude, meaning, self-acceptance, and valued identity. A comparative group study in depressive disorders found that positive psychotherapy and CBT both produced clinical benefits, supporting the feasibility of combining their techniques when the treatment rationale is coherent.[23] Compassion-oriented work similarly suggests that shame and perfectionism may soften when participants develop a less punitive stance toward failure.[24] In a CBT-P sequence, positive exercises are not decorative additions; they can supply an alternative source of self-evaluation after rigid achievement rules are questioned.

Group delivery may be particularly appropriate for university settings. It allows participants to observe that peers also conceal mistakes, revise goals, and survive imperfect performance. Such interpersonal learning can weaken the assumption that everyone else meets standards effortlessly. It also permits repeated behavioural practice within a socially relevant context. From a service perspective, group treatment can expand capacity where counselling centres face high demand. Meta-analytic evidence for group CBT in depression indicates that the format can be both efficacious and acceptable.[25] Nonetheless, group averages alone do not establish whether a perfectionism intervention changed the intended processes. The important question is whether improvement concentrated in mistake concern and action doubt, or whether all subscales shifted indiscriminately.

Outcome selectivity is therefore central to interpretation. Personal standards may remain stable because they can reflect values, identity, or adaptive striving. Parental expectations and criticism may also be comparatively resistant to a student-only intervention because they refer partly to enduring relational histories. By contrast, concern over mistakes and doubts about actions are proximal cognitive processes that can be practised during treatment. Neurocognitive and self-report work has supported the distinction between positive and negative perfectionism dimensions,[27] while bifactor modelling has shown that common variance and dimension-specific variance must be disentangled when interpreting perfectionism scores.[28] Functional comparisons of adaptive and maladaptive perfectionists likewise indicate that similar outward standards can coexist with very different compulsive behaviour and distress.[29] A selective change pattern would thus strengthen construct validity: improvement should be largest in the components the sessions directly rehearsed.

Durability adds a second requirement. An immediate fall in symptom scores may reflect group support, expectancy, attention, or temporary relief after structured sessions end. Maintenance after eight weeks is more consistent with retained cognitive and behavioural learning, although it does not establish long-term remission. The trajectory between post-treatment and follow-up is particularly informative. Some rebound is plausible because students again encounter deadlines and evaluation without weekly therapist contact. A durable intervention need not preserve 100% of its immediate gain; rather, it should retain a substantial portion while continuing to separate from an untreated comparison group. Depression and anxiety may also follow different courses from perfectionism because affective symptoms respond to changes in activity, hope, and social connection as well as to cognitive appraisal.

The analysis of controlled longitudinal studies requires more than a sequence of within-group significance tests. Pretreatment comparability does not justify treating one arm's change as causal while ignoring the other arm. Between-arm contrasts at follow-up, difference-in-differences, and standardized effects answer related but distinct questions. Methodological guidance recommends analysing follow-up outcomes in relation to pretreatment values and reporting effect sizes with uncertainty rather than relying only on p values.[31] Change-score and covariance approaches can differ according to reliability, imbalance, and model assumptions.[33,34] Pretest–posttest effect-size methods further show that the numerator and standardizer must be defined transparently.[32] These considerations motivated an analysis that separates magnitude, selectivity, and persistence without claiming unavailable participant-level precision.

The present study asked whether an eight-session CBT-P group intervention produced an outcome-specific benefit that remained evident eight weeks later. Three expectations followed from the treatment content. First, concern over mistakes, doubts about actions, the maladaptive FMPS composite, depressive symptoms, and anxiety symptoms were expected to show moderate or larger separation from the waitlist immediately after treatment. Second, a substantial majority of the differential improvement in these outcomes was expected to remain at follow-up. Third, personal standards, parental expectations, and parental criticism were expected to show smaller separation because they were either potentially adaptive or less directly modifiable through

student-only sessions. By testing these expectations together, the study moves from a binary efficacy question to a clinically sharper account of what changed, how strongly it changed, and how much of that change persisted.

2. METHODS

The numerical record comprised arm-level counts, means, standard deviations, test statistics, and effect estimates from a randomized waitlist-controlled trial of group CBT-P for Chinese university students.[1] Ninety-six students entered screening, 80 met score requirements, 72 completed a face-to-face eligibility interview, and 64 were allocated in a sex-stratified 1:1 ratio to CBT-P or waitlist conditions. Complete observations at pretreatment (T1), post-treatment (T2), and eight-week follow-up (T3) were available for 27 CBT-P participants and 29 waitlist participants. The analytical sample therefore contained 56 students.

The comparison was longitudinal and outcome-focused. T1 represented the assessment before the first session, T2 occurred immediately after the eighth weekly session, and T3 occurred eight weeks after T2. Lower scores indicated improvement on all analysed outcomes. The principal inferential evidence consisted of the reported time effects and time-by-group interactions. Additional calculations were performed directly from the arm-level means and standard deviations. No participant-level covariance matrix was available; accordingly, no within-person standard error, multilevel model, mediation coefficient, or individual response classification was inferred.

2.1. Participants and allocation

Eligible participants were undergraduate or postgraduate students willing to attend the group programme. Entry required an FMPS total score of at least 84 and a Self-Rating Depression Scale (SDS) standard score from 53 to 72 or a Self-Rating Anxiety Scale (SAS) standard score from 50 to 70. Students were excluded for hallucinations or other acute psychotic episodes, immediate risk of suicide or injury, concurrent psychological treatment, or a change in psychotropic medication during the preceding three months. Random numbers were used separately within female and male groups, yielding 32 participants per condition before attrition.

The complete-case sample had a mean age of 21.1 years (SD 2.2); 34 participants were women and 22 were men. Thirty-five were bachelor's students and 21 were master's students. The analysed CBT-P and waitlist groups were similar in sex distribution, education level, marital status, previous psychological treatment, and previous psychotropic medication. Thirty CBT-P participants attended all eight sessions, and 27 completed every assessment. The waitlist condition retained 29 complete cases.

2.2. Intervention

CBT-P was delivered in eight weekly 120-minute group sessions by a licensed counsellor with more than 15 years of clinical and group-work experience, assisted by three graduate trainees in psychology or social work. The sequence began with identification of perfectionism-related behaviours and their consequences. Participants then learned to recognise automatic thoughts, connect beliefs to emotional and behavioural responses, and test the assumption that errors necessarily produce rejection or failure. Behavioural experiments and cognitive restructuring challenged inflexible rules. A dedicated session addressed procrastination through immediate-action planning and graded task engagement.

The latter half of the programme incorporated positive psychotherapy. Participants reconsidered past mistakes from a constructive perspective, practised optimism and process-oriented definitions of success, examined family influences on perfectionism, reconstructed core beliefs that tied worth to flawless performance, identified strengths, and rehearsed confident behaviour. The final session consolidated a self-description that could acknowledge ambition without equating imperfection with personal inadequacy. This sequencing preserved the behavioural specificity of CBT while using positive exercises to strengthen self-acceptance and motivation.

The waitlist condition received no intervention during the 16-week observation period. Assessments for CBT-P participants were completed face to face at the university mental health centre, whereas waitlist assessments were completed online at scheduled times with reminders. Written consent was obtained, participation was voluntary, and participants could withdraw. Personal identifiers were removed from the electronic records.

2.3. Measures

The FMPS contains 35 items rated from 1 to 5. Concern over mistakes (CM) assesses distress and negative interpretation following error; doubts about actions (DA) assesses uncertainty about whether tasks have been completed adequately; personal standards (PS) assesses demanding self-set expectations; parental expectations (PE) and parental criticism (PC) assess perceived family demands and disapproval. The maladaptive perfectionism composite (MFMPS) summed CM, DA, PE, and PC. Internal-consistency coefficients for the translated subscales in the analysed sample ranged from 0.74 to 0.91.

The SDS contains 20 items covering depressive experiences during the previous week. Standard scores below 53 indicate no depression, 53–62 mild depression, 63–72 moderate depression, and 73 or above severe depression under the Chinese norm. Internal consistency was $\alpha = 0.70$. The 20-item SAS measures subjective anxiety; standard scores below 50 indicate no anxiety, 50–59 mild anxiety, 60–70 moderate anxiety, and 71 or above severe anxiety. Internal consistency was $\alpha = 0.72$. These reliability values are adequate for group comparisons but add

uncertainty to fine-grained classification of individuals.

2.4. Analytical strategy

The first layer reproduced the longitudinal arm profiles for all eight outcomes and retained the reported repeated-measures time and time-by-group tests. The second layer quantified standardized arm separation at T2 and T3. For outcome k at time t, Cohen's standardized difference was

$$d_{kt} = \frac{\bar{Y}_{\mathrm{CBT-P},kt} - \bar{Y}_{\mathrm{WL},kt}}{S_{p,kt}}, \tag{1}$$

where $S_{p,kt}$ was the pooled cross-sectional standard deviation. Because the analysed arms were modest in size, d_{kt} was multiplied by the correction

$$J = 1 - \frac{3}{4(n_{\mathrm{CBT-P}} + n_{\mathrm{WL}}) - 9} \tag{2}$$

to obtain Hedges' g_{kt}. Negative values denoted lower scores in CBT-P. For readability, figures displaying intervention advantage reversed the sign, so positive values denoted benefit. Approximate 95% compatibility intervals used

$$\mathrm{SE}(g) = \sqrt{\frac{n_1 + n_0}{n_1 n_0} + \frac{g^2}{2(n_1 + n_0 - 2)}}. \tag{3}$$

These intervals describe cross-sectional arm separation under an independent-groups approximation. They do not replace a repeated-measures model. Effect-size reporting followed recommendations to distinguish magnitude from dichotomous significance decisions.[35–37]

The third layer measured differential improvement on each scale:

$$\begin{aligned} B_{k,t} = {} & \left(\bar{Y}_{\mathrm{CBT-P},k,T1} - \bar{Y}_{\mathrm{CBT-P},k,t}\right) \\ & - \left(\bar{Y}_{\mathrm{WL},k,T1} - \bar{Y}_{\mathrm{WL},k,t}\right), \quad t \in \{T2, T3\}. \end{aligned} \tag{4}$$

Positive $B_{k,t}$ values indicate that CBT-P improved more than the waitlist in the original units of the scale. Persistence was expressed as

$$R_k = \frac{B_{k,T3}}{B_{k,T2}}. \tag{5}$$

The coefficient was interpreted only for the five outcomes with a significant time-by-group interaction and a positive T2 differential improvement. A value of 1.00 indicates complete retention of the T2 advantage, a value between 0 and 1 indicates partial retention, and a value above 1 indicates a larger T3 advantage. Ratios were not interpreted for near-zero T2 changes because small denominators make them unstable.

The fourth layer tested selectivity. The responsive set comprised CM, DA, MFMPS, SDS, and SAS. The comparatively stable set comprised PS, PE, and PC. At each post-randomization assessment, the endpoint-specificity ratio divided the mean absolute Hedges' g in the responsive set by the mean absolute g in the stable set. A ratio substantially above 1 indicates concentration of separation in the intended domains. Robustness was then assessed by moving the criterion for a notable standardized advantage through 0.50, 0.60, 0.70, and 0.80. This perturbation test was descriptive; it examined whether the conclusion depended on a single conventional cutoff.

No missing values were imputed because the required person-level patterns were unavailable. This choice avoids creating artificial precision, but it means the calculations describe completers rather than a strict intention-to-treat population. Contemporary guidance recommends prespecifying missing-data assumptions and using all randomized participants where possible.[40–42] The consequence of complete-case analysis is addressed directly in the limitations.

3. RESULTS

3.1. Participation and pretreatment profile

Participant progression is displayed in figure 1. The largest reduction occurred between enrolment and score eligibility, where 16 of 96 students did not meet the required symptom profile. Eight additional students did not complete the eligibility interview, and eight interviewed students were not randomized because of schedule changes or unmet criteria. Of 64 randomized students, 56 provided all three assessments, corresponding to complete-assessment retention of 87.5%. Retention differed modestly by arm: 84.4% in CBT-P and 90.6% in the waitlist. This six-percentage-point difference is small in count terms but remains relevant because complete-case estimates can be biased if noncompletion is related to outcome.

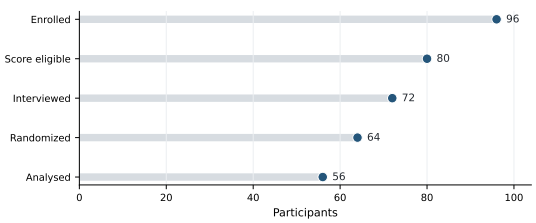

Figure 1. Participant progression.

The count profile shows that most losses occurred before random allocation; only eight of the 64 randomized students were absent from the complete-assessment analysis.

The analysed groups were demographically comparable (table 1). Women formed approximately three fifths of each arm, and the proportions of bachelor's and master's students were similar. Previous psychological treatment

and psychotropic medication were uncommon in both conditions. Age differed by 1.4 years in the displayed arm means, although the tabulated waitlist mean of 20.0 years is not fully consistent with the combined mean of 21.1 years and group sizes. The age entry was therefore preserved as reported rather than silently recalculated. This minor inconsistency reinforces the need for participant-level verification in future confirmatory work but does not affect the outcome calculations.

The demographic distribution supports comparability of the analysed arms, although the small sample cannot exclude imbalance in unrecorded clinical or academic characteristics.

At T1, the two arms differed little on the eight outcomes. Absolute standardized differences calculated from the displayed summaries ranged from 0.06 to 0.19, which is compatible with successful random allocation in a modest sample. The combined SDS mean of 60.50 (SD 4.80) fell within the mild-depression interval of the stated Chinese norm, whereas the combined SAS mean of 62.34 (SD 5.80) fell within the moderate-anxiety interval. The FMPS total mean of 109.91 (SD 13.41) was consistent with pronounced perfectionistic difficulty under the study's entry rule. These values define a symptomatic student sample rather than a formally diagnosed psychiatric population.

3.2. Longitudinal outcome profiles

Arm means and interaction tests are shown in table 2. Time-by-group interactions were large for CM ($\eta^2 = 0.38$), MFMPS ($\eta^2 = 0.47$), and SAS ($\eta^2 = 0.51$); they were also evident for DA ($\eta^2 = 0.18$) and SDS ($\eta^2 = 0.21$). PS, PE, and PC had small interaction estimates and nonsignificant tests. The pattern is more informative than a statement that five tests crossed a threshold: every dimension that directly represented maladaptive appraisal or affect showed separation, whereas personal standards and the two parental-evaluation dimensions did not.

The joint pattern of means and interactions indicates selective movement in maladaptive appraisal and affect rather than a uniform decrease across all perfectionism dimensions.

The MFMPS trajectory in figure 2 clarifies the composite result. The CBT-P mean declined by 8.00 points from T1 to T2, whereas the waitlist mean increased by 1.00 point, producing a 9.00-point differential improvement. At T3, the CBT-P mean had risen by 1.71 points from its post-treatment low but remained 6.29 points below T1. The waitlist remained slightly above T1. The resulting T3 differential improvement was 7.11 points. Thus, the follow-up pattern indicates partial attenuation rather than disappearance: nearly four fifths of the immediate composite advantage remained.

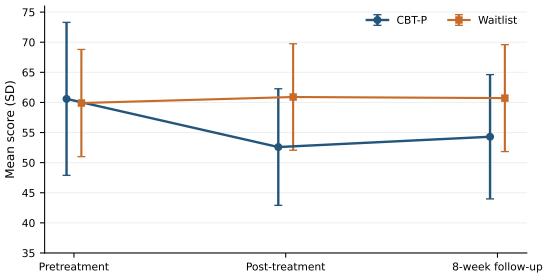

Figure 2. Maladaptive perfectionism trajectory.

The plotted means show a large immediate composite decline, followed by modest rebound while a clear arm difference remains.

Depressive symptoms followed a related but not identical course (figure 3). The CBT-P mean fell 4.52 points by T2 and remained 3.46 points below T1 at T3. The waitlist mean declined by 1.04 points at T2 but returned close to T1 at follow-up. Consequently, the between-arm standardized advantage was slightly larger at T3 than at T2 even though CBT-P participants showed some rebound. This distinction illustrates why maintenance should not be judged solely from the treated arm: the comparative effect depends on both trajectories. The retained differential improvement for SDS was 88.2%, the highest among the five responsive endpoints.

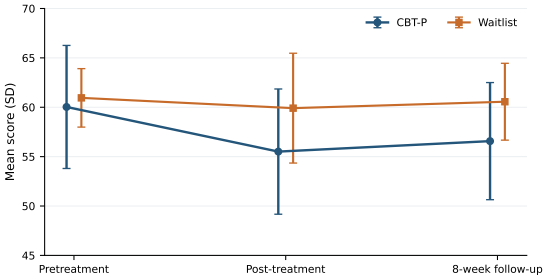

Figure 3. Depressive symptom trajectory.

The symptom profile indicates that the relative depression advantage was preserved because the waitlist returned toward its pretreatment level while most of the CBT-P reduction remained.

Anxiety scores are displayed in figure 4. The CBT-P mean declined 5.47 points by T2 and remained 4.45 points below T1 at T3. Waitlist means were essentially flat across the three assessments. The immediate differential improvement was therefore 5.85 points and the follow-up differential improvement was 4.49 points, corresponding to 76.8% retention. The treated mean moved from the moderate-anxiety interval into the mild-anxiety interval at both later assessments, while the waitlist mean remained in the moderate interval. Because these are arm means rather than individual classifications, this norm-referenced shift should be understood as a change in group location, not a remission rate.

Table 1. Characteristics of the analysed sample.

Characteristic	Overall ($N = 56$)	CBT-P ($n = 27$)	Waitlist ($n = 29$)
Age, mean (SD), years	21.1 (2.2)	21.4 (2.3)	20.0 (2.1)
Men, n (%)	22 (39.3)	11 (40.7)	11 (37.9)
Women, n (%)	34 (60.7)	16 (59.3)	18 (62.1)
Bachelor's level, n (%)	35 (62.5)	16 (59.3)	19 (65.5)
Master's level, n (%)	21 (37.5)	11 (40.7)	10 (34.5)
Single, n (%)	45 (80.4)	22 (81.5)	23 (79.3)
Married/partnered, n (%)	11 (19.6)	5 (18.5)	6 (20.7)
Previous psychological treatment, n (%)	8 (14.3)	3 (11.1)	5 (17.2)
Previous psychotropic medication, n (%)	5 (8.9)	2 (7.4)	3 (10.3)

Percentages are within columns. CBT-P, cognitive behavioural therapy with positive psychotherapy.

Table 2. Outcome profiles across the three assessments.

Outcome	Arm	T1	T2	T3	Time × group: F (p; η^2)
CM	CBT-P	21.74 (7.55)	17.00 (4.88)	17.81 (5.97)	16.28 ($< .001$; .38)
	Waitlist	21.10 (5.19)	21.76 (4.99)	21.21 (4.88)	
PS	CBT-P	24.07 (4.85)	23.78 (4.20)	23.56 (4.52)	1.74 (.186; .02)
	Waitlist	23.62 (3.52)	23.03 (4.36)	23.48 (3.72)	
PE	CBT-P	15.52 (4.00)	15.07 (4.13)	15.11 (4.10)	1.74 (.185; .06)
	Waitlist	16.28 (4.36)	16.00 (3.71)	16.34 (3.79)	
PC	CBT-P	9.52 (2.91)	9.19 (3.10)	9.37 (2.86)	1.59 (.214; .06)
	Waitlist	9.24 (2.29)	9.62 (1.72)	9.62 (1.59)	
DA	CBT-P	13.81 (3.76)	11.33 (2.90)	12.00 (2.87)	5.75 (.005; .18)
	Waitlist	13.28 (2.53)	13.52 (2.89)	13.55 (2.43)	
MFMPS	CBT-P	60.59 (12.70)	52.59 (9.68)	54.30 (10.32)	23.78 ($< .001$; .47)
	Waitlist	59.90 (8.90)	60.90 (8.84)	60.72 (8.88)	
SDS	CBT-P	60.03 (6.23)	55.51 (6.34)	56.57 (5.93)	6.99 (.002; .21)
	Waitlist	60.95 (2.96)	59.91 (5.56)	60.56 (3.89)	
SAS	CBT-P	62.69 (5.86)	57.22 (6.36)	58.24 (6.89)	27.61 ($< .001$; .51)
	Waitlist	62.03 (5.84)	62.41 (7.00)	62.07 (5.60)	

Values at T1–T3 are mean (SD). CM, concern over mistakes; PS, personal standards; PE, parental expectations; PC, parental criticism; DA, doubts about actions; MFMPS, maladaptive FMPS composite; SDS, depressive symptoms; SAS, anxiety symptoms.

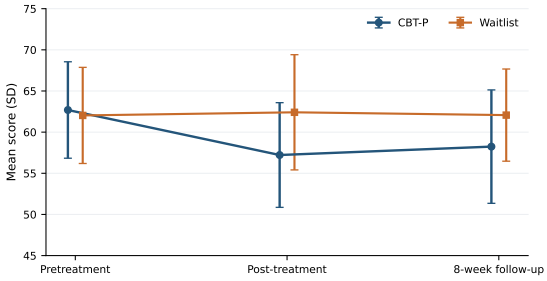

Figure 4. Anxiety symptom trajectory.

The near-horizontal waitlist line strengthens the interpretation that the anxiety separation arose from improvement in CBT-P rather than deterioration in the comparison arm.

3.3. Standardized separation and endpoint specificity

Small-sample-corrected estimates are presented in table 3. At T2, Hedges' g ranged from -0.73 to -0.95 across the responsive endpoints, indicating moderate-to-large lower scores in CBT-P. All five approximate intervals excluded zero. At T3, values ranged from -0.58 to -0.79, and their intervals also excluded zero under the independent-groups approximation. In contrast, PS, PE, and PC estimates remained small at both assessments, and all corresponding intervals included zero. The distinction persisted despite the reduction in magnitude for several responsive outcomes.

The corrected effect estimates confirm that every responsive endpoint retained at least a moderate arm difference, while the other dimensions remained imprecise

Table 3. Standardized separation, differential improvement, and retention.

Outcome	T2 g (95% CI)	T3 g (95% CI)	B_{T2}	B_{T3}	Retention (%)
CM	−0.95 (−1.51, −0.40)	−0.62 (−1.15, −0.08)	5.40	4.04	74.8
DA	−0.75 (−1.29, −0.20)	−0.58 (−1.11, −0.04)	2.72	2.08	76.5
MFMPS	−0.89 (−1.44, −0.34)	−0.66 (−1.20, −0.12)	9.00	7.11	79.0
SDS	−0.73 (−1.27, −0.19)	−0.79 (−1.34, −0.25)	3.48	3.07	88.2
SAS	−0.76 (−1.31, −0.22)	−0.60 (−1.14, −0.07)	5.85	4.49	76.8
PS	0.17 (−0.35, 0.70)	0.02 (−0.51, 0.54)	−0.30	0.37	–
PE	−0.23 (−0.76, 0.29)	−0.31 (−0.84, 0.22)	0.17	0.47	–
PC	−0.17 (−0.70, 0.35)	−0.11 (−0.63, 0.42)	0.71	0.53	–

Negative g indicates a lower CBT-P mean. B is the arm difference in improvement in the scale's original units. Retention was interpreted only for outcomes with a positive T2 differential improvement and significant interaction. Intervals use an independent-groups approximation.

and small.

The paired standardized profile in figure 5 makes the selectivity result visible without collapsing distinct constructs into a single total. Five outcomes lie to the right of the 0.50 reference at both T2 and T3. The other three cluster near zero and never cross the reference. Mean absolute g among responsive endpoints was 0.815 at T2 and 0.649 at T3. Corresponding means in the comparatively stable set were 0.193 and 0.145, producing specificity ratios of 4.22 and 4.48. The increase in the ratio at follow-up does not mean all responsive effects grew; it means that their separation remained much larger relative to dimensions that were not expected to move.

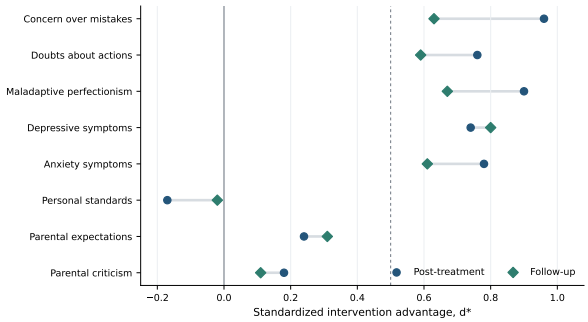

Figure 5. Endpoint-specific intervention advantage.

The visual separation between the two endpoint groups is preserved at both assessments, supporting treatment specificity despite attenuation in several responsive outcomes.

Threshold perturbation yielded the same substantive ordering (table 4). All five responsive endpoints exceeded 0.50 at T2 and T3. Raising the criterion to 0.60 retained all five at T2 and four at T3; at 0.70 all five remained at T2, but only SDS exceeded the criterion at T3. A threshold of 0.80 retained two immediate effects and none at follow-up. No PS, PE, or PC estimate exceeded even the lowest criterion. The result is therefore robust at a conventional moderate-effect definition, while claims of uniformly large long-term effects would be too strong.

The threshold counts show that the moderate-effect conclusion is stable, whereas a claim of uniformly large persistence is not supported.

3.4. Persistence of differential improvement

Retention coefficients for the five responsive outcomes are compared in figure 6. All fell within a narrow 74.8–88.2% range, with a mean of 79.1%. CM showed the greatest attenuation, retaining 74.8% of its T2 differential improvement; SDS showed the least attenuation, retaining 88.2%. DA, MFMPS, and SAS each retained approximately three quarters to four fifths. A domain-balanced standardized index, defined as the mean of the MFMPS effect and the average SDS/SAS effect, decreased from 0.816 at T2 to 0.678 at T3, retaining 83.1% of its immediate magnitude. The agreement between the raw-change and standardized summaries indicates that persistence was not an artefact of one scale's numerical range.

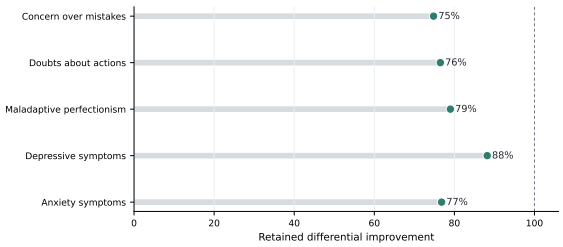

Figure 6. Retention of differential improvement.

The compact range of retention coefficients indicates a broadly shared maintenance pattern, with depressive symptoms preserving the largest proportion of differential improvement.

The pattern also distinguishes comparative persistence from complete symptom stability. CBT-P means rose between T2 and T3 for CM, DA, MFMPS, SDS, and SAS, indicating some loss of the immediate gain. Waitlist means, however, showed no parallel sustained improvement. The appropriate conclusion is therefore neither "no relapse" nor "complete maintenance." A more accurate statement is that most of the arm difference created

Table 4. Threshold perturbation of standardized advantage.

Criterion	Responsive T2	Responsive T3	Stable T2	Stable T3
$\lvert g\rvert \geq 0.50$	5/5	5/5	0/3	0/3
$\lvert g\rvert \geq 0.60$	5/5	4/5	0/3	0/3
$\lvert g\rvert \geq 0.70$	5/5	1/5	0/3	0/3
$\lvert g\rvert \geq 0.80$	2/5	0/5	0/3	0/3

Responsive outcomes: CM, DA, MFMPS, SDS, SAS. Stable dimensions: PS, PE, PC.

during treatment remained eight weeks later, although perfectionism-related cognition and affective symptoms showed modest rebound after weekly sessions ended.

4. DISCUSSION

The research question concerned selectivity and persistence: did CBT-P preferentially change maladaptive appraisal and affect, and did that pattern remain after treatment? The answer is yes within the limits of the available aggregate record. Concern over mistakes, doubts about actions, maladaptive perfectionism, depression, and anxiety all showed moderate or larger standardized separation immediately after treatment and retained advantages eight weeks later. Personal standards, parental expectations, and parental criticism remained comparatively stable. The effect was not an indiscriminate downward shift across self-report scales. Instead, change concentrated in processes that the sessions directly addressed: punitive interpretation of mistakes, uncertainty about action, procrastination, rigid self-evaluation, and the affective burden associated with those processes.

This selectivity is clinically meaningful. If every FMPS dimension had fallen by a similar amount, it would be difficult to know whether participants had developed healthier flexibility or had merely become less willing to endorse demanding statements. Stable PS scores suggest that students did not need to abandon aspiration in order to improve. That result aligns with the broader distinction between perfectionistic strivings and perfectionistic concerns.[4,5] High standards can support valued achievement when they are flexible, domain-limited, and not used as the sole measure of worth. Treatment may therefore be framed as increasing choice around standards rather than lowering them. Such language is especially important in academic cultures where effort, family obligation, and achievement can carry positive social meaning.

The CM and DA changes fit the behavioural content of the intervention. Concern over mistakes can decline when participants intentionally produce an adequate rather than flawless performance, observe the actual consequences, and compare them with catastrophic predictions. Doubt about action can decline when repeated checking and procrastination are replaced by completion rules and toleration of uncertainty. The immediate standardized advantages for CM and DA were 0.95 and 0.75, respectively. Their retained differential improvements of 74.8% and 76.5% indicate that learning remained evident but was not fully stable. Continued exposure to mistakes and uncertainty may be required after group sessions end. Booster meetings, brief digital prompts, or peer practice could test whether the remaining quarter of the immediate advantage can be preserved without turning maintenance activities into new perfectionistic obligations.

The absence of clear change in PE and PC is also theoretically coherent. These subscales ask about perceived parental demands and criticism, which may reflect longstanding relational memories rather than only current cognitive habits. The programme helped students examine family influences, but parents did not participate. A student can reinterpret a parent's expectation without changing the historical fact that the expectation was communicated. Measurement content may therefore limit how much these scores can move over eight or sixteen weeks. Future work should distinguish perceived current pressure from remembered upbringing and should test whether optional family sessions alter present relational processes. Such additions would require careful cultural adaptation; parental involvement should not presume that family expectations are uniformly harmful.

The symptom findings support a transdiagnostic interpretation. Depression and anxiety improved alongside maladaptive perfectionism even though the intervention was organised around perfectionistic processes rather than separate disorder protocols. Perfectionism has been proposed as a treatment barrier because fear of failure can interfere with homework, exposure, decision making, and willingness to tolerate incomplete progress. Research with clinically anxious children found that perfectionism could influence CBT outcomes,[26] and broader reviews have linked perfectionistic concerns to diverse disorders.[6] Reducing error-related threat may therefore release behavioural engagement across several symptom domains. The present pattern is compatible with that account, although it cannot prove that perfectionism change mediated symptom change because the necessary person-level temporal covariances are absent. Mediation requires evidence that treatment changes the mediator, that mediator change precedes or predicts outcome change, and that alternative explanations are sufficiently controlled.[43]

Depression showed the strongest persistence. Its comparative standardized effect increased from 0.73 to 0.79, and 88.2% of the raw differential improvement remained. This does not mean depressive symptoms continued to decline in the CBT-P arm; the arm mean rose modestly after treatment. The larger comparative effect arose because the waitlist returned close to its pretreatment level. Positive psychotherapy components may have contributed to this relative stability by broadening self-evaluation beyond failure and productivity. Strength identification, hope-focused practice, and a process-oriented definition of success can support activity and self-acceptance even when perfectionistic thoughts recur. The interpretation remains tentative because the design did not isolate positive psychotherapy from CBT. A dismantling trial would be needed to estimate the incremental contribution of these components.

Anxiety retained 76.8% of its differential improvement, and the CBT-P arm mean remained within the mild rather than moderate range at follow-up. Anxiety may be particularly sensitive to renewed evaluation and uncertainty after structured support ends. Participants returning to examinations, presentations, or deadlines would again encounter triggers that activate checking and avoidance. Maintenance planning should therefore include repeated practice under real academic pressure, not only discussion in the final session. Ecological momentary assessments could capture short-lived spikes in threat and determine whether the intervention changes recovery time after mistakes. Such a design would add temporal resolution without assuming that a single questionnaire score represents the whole eight-week interval.

The specificity ratio provides a useful complement to individual effect sizes. Responsive outcomes showed more than four times the average standardized separation of the stable dimensions at both post-treatment and follow-up. This ratio is not a universal clinical index, and it should not be compared across studies without harmonised outcome sets. Its value here is internal: it tests whether the response pattern corresponds to the intervention's conceptual target. The threshold perturbation reaches the same conclusion. Every responsive endpoint exceeded 0.50 at both assessments, while none of the stable dimensions did. At stricter criteria, follow-up effects were no longer uniformly large. This graded account is preferable to a binary description based only on $p < .05$. Statistical statements should be compatible with effect magnitude, uncertainty, design quality, and clinical context.[38,39]

The findings also speak to intervention efficiency. Eight two-hour sessions represent 16 therapist-led hours per group, with several students treated simultaneously. Group delivery may allow university counselling services to reach more students than individual weekly therapy, particularly when therapist availability is limited. Efficiency, however, cannot be established from contact hours alone. Preparation, screening, assistant time, room availability, missed sessions, supervision, and post-group support all contribute to cost. Formal health-economic evaluation should estimate cost per reliable improvement and compare group CBT-P with individual CBT, guided digital treatment, active psychoeducation, and stepped care. Existing evidence indicates that group CBT can be effective for depression,[25] but local service costs and student preferences will determine whether CBT-P is the most efficient option.

Several design features strengthen the interpretation. Random allocation reduced systematic pretreatment differences, and the T1 standardized contrasts were uniformly small. Measures covered both the proposed cognitive target and associated affective symptoms. The eight-week follow-up permitted examination of persistence rather than immediate response alone. Reporting both arm trajectories prevented improvement from being attributed to treatment merely because one within-arm test was significant. Small-sample correction reduced upward bias in standardized estimates. Finally, the selectivity and perturbation analyses asked whether the result was coherent across constructs and thresholds rather than relying on one favourable outcome.

Important limitations remain. First, the analytical sample contained 56 completers from 64 randomized students. If withdrawal was related to symptom burden, treatment dissatisfaction, or schedule conflict associated with stress, complete-case effects could be biased. The slightly lower retention in CBT-P makes this concern relevant even though the count difference was small. A future trial should preserve every randomized participant in the main analysis, document reasons for noncompletion by arm, and use multiple imputation or likelihood-based longitudinal modelling under clearly stated assumptions. Sensitivity analyses should then examine departures from missing-at-random assumptions.

Second, the arm-level record does not contain correlations among repeated observations. The present compatibility intervals therefore quantify cross-sectional separation and cannot recover the precision of change scores. Nor can they identify individual response, reliable change, mediation, moderators, or latent trajectory classes. Individual records would permit mixed-effects models with time treated categorically, treatment-by-time contrasts, robust standard errors, and prespecified covariate adjustment. They would also permit estimation of the association between change in CM or DA and subsequent change in SDS or SAS. Until those analyses are available, the mechanism remains plausible rather than demonstrated.

Third, the waitlist controls for passage of time but not for attention, expectancy, group contact, or demand characteristics. Participants knew whether they were receiving treatment, and assessment mode differed between arms. Face-to-face completion in CBT-P and online completion in the waitlist could introduce mode effects. An active comparison group receiving structured study-skills

support or general stress management would better isolate the contribution of perfectionism-focused CBT-P. Blinded outcome assessment is difficult with self-report, but independent diagnostic interviews and behavioural tasks could reduce reliance on treatment-congruent questionnaires.

Fourth, participants were volunteers selected by elevated FMPS and symptom scores rather than by a psychiatric diagnosis. Results apply most directly to symptomatic university students motivated to join a group. They should not be generalized automatically to severe depression, acute suicide risk, psychosis, or students receiving concurrent therapy. The SDS and SAS are screening instruments, and their norm categories do not establish clinical diagnoses. Future trials should report diagnostic status, comorbidity, medication stability, and adverse events, while retaining a clear pathway for crisis referral.

Fifth, follow-up lasted eight weeks. The retention coefficients show short-term persistence, not long-term prevention. Academic stress varies across semesters, and an assessment taken during a quiet period may underestimate relapse under examination pressure. Measurements at six and twelve months, aligned with the academic calendar, would reveal whether the response survives repeated performance demands. Treatment fidelity also requires stronger documentation. Session checklists, independent ratings of recorded sessions, therapist competence measures, and homework completion would clarify whether observed variation reflects programme content or delivery.

Sixth, several outcomes overlap conceptually and statistically. MFMPS includes CM, DA, PE, and PC, so treating the composite and its components as independent evidence would overstate multiplicity. The analysis avoided formal combination of their p values and used the specificity ratio descriptively. A future confirmatory protocol should define one primary perfectionism endpoint, one principal symptom endpoint, and a limited set of mechanistic secondary outcomes. CONSORT reporting and prospective registration would improve transparency about allocation, attrition, outcome hierarchy, and deviations.[30]

Clinical implementation should retain the treatment's selective emphasis. Therapists can validate ambition while questioning the rule that errors determine worth. Behavioural practice should be concrete: submit an assignment after a predetermined number of checks, make a low-stakes contribution without exhaustive preparation, or complete a task to an agreed "good enough" criterion. Positive psychotherapy can then consolidate a broader identity by linking strengths and values to actions that are not contingent on perfect performance. Progress should be reviewed using separate measures of evaluative concern and standards so that improvement is not misread as loss of motivation.

Research design can also be sharpened. A three-arm trial comparing CBT, CBT-P, and an active student-support condition would determine whether positive psychotherapy adds durable benefit beyond CBT alone. Repeated assessments during treatment could establish temporal ordering. A preregistered primary contrast could test MFMPS at post-treatment, with SDS and SAS as key secondary outcomes. Person-level mixed models should include all randomized students, while sensitivity analyses evaluate missingness and therapist clustering. Qualitative interviews could examine whether participants experienced the intervention as reduced self-attack, increased flexibility, broader self-worth, or simply temporary reassurance. This combination would connect statistical change to lived mechanisms without sacrificing inferential discipline.

5. CONCLUSION

The paper asked whether group CBT-P produced a selective and persistent response rather than an indiscriminate reduction across perfectionism scales. The observed pattern answers that question affirmatively. Immediately after eight sessions, concern over mistakes, doubts about actions, maladaptive perfectionism, depression, and anxiety showed standardized intervention advantages of 0.73–0.95; eight weeks later, advantages of 0.58–0.79 remained. On average, 79.1% of differential improvement was retained. Personal standards, parental expectations, and parental criticism stayed comparatively stable, and the responsive domains showed more than four times their average separation. CBT-P therefore appears to reduce punitive error processing and associated distress without requiring students to relinquish high aspirations. The conclusion is limited to short-term complete-case, aggregate evidence; a larger active-controlled trial with participant-level longitudinal analysis is required to establish long-term durability and mechanism.

FUNDING

No external funding was received for this work.

CONFLICT OF INTEREST

The author declares no commercial or financial relationships that could be construed as a potential conflict of interest.

BIBLIOGRAFIA / REFERENCES

1. Zuo Z, Zhang X. A randomized controlled trial of group CBT with positive psychotherapy intervention for university students with maladaptive perfectionism in China. *Front Psychol*. 2023;14:1161575. doi:10.3389/fpsyg.2023.1161575.
2. Frost RO, Marten P, Lahart C, Rosenblate R. The dimensions of perfectionism. *Cogn Ther Res*. 1990;14:449–468. doi:10.1007/BF01172967.
3. Cheng SK, Chong GH, Wong CW. Chinese Frost Multidimensional Perfectionism Scale:

a validation and prediction of self-esteem and psychological distress. *J Clin Psychol*. 1999;55:1051–1061. doi:10.1002/(SICI)1097-4679(199909)55:9<1051::AID-JCLP3>3.0.CO;2-1.
4. Stoeber J, Otto K. Positive conceptions of perfectionism: approaches, evidence, challenges. *Pers Soc Psychol Rev*. 2006;10:295–319. doi:10.1207/s15327957pspr1004_2.
5. Bieling PJ, Israeli AL, Antony MM. Is perfectionism good, bad, or both? Examining models of the perfectionism construct. *Pers Individ Dif*. 2004;36:1373–1385. doi:10.1016/S0191-8869(03)00235-6.
6. Limburg K, Watson HJ, Hagger MS, Egan SJ. The relationship between perfectionism and psychopathology: a meta-analysis. *J Clin Psychol*. 2017;73:1301–1326. doi:10.1002/jclp.22435.
7. Egan SJ, Wade TD, Shafran R. Perfectionism as a transdiagnostic process: a clinical review. *Clin Psychol Rev*. 2011;31:203–212. doi:10.1016/j.cpr.2010.04.009.
8. Hill AP, Curran T. Multidimensional perfectionism and burnout: a meta-analysis. *Pers Soc Psychol Rev*. 2016;20:269–288. doi:10.1177/1088868315596286.
9. Fong RW, Yuen M. Perfectionism in Chinese elementary school students: validation of the Chinese Adaptive/Maladaptive Perfectionism Scale. *Talent Dev Excell*. 2011;3:203–213.
10. Wang KT, Slaney RB, Rice KG. Perfectionism in Chinese university students from Taiwan: a study of psychological well-being and achievement motivation. *Pers Individ Dif*. 2007;42:1279–1290. doi:10.1016/j.paid.2006.10.006.
11. DiBartolo PM, Rendón MJ. A critical examination of the construct of perfectionism and its relationship to mental health in Asian and African Americans using a cross-cultural perspective. *Clin Psychol Rev*. 2012;32:139–152. doi:10.1016/j.cpr.2011.09.007.
12. Grzegorek JL, Slaney RB, Franze S, Rice KG. Self-criticism, dependency, self-esteem, and grade point average satisfaction among clusters of perfectionists and nonperfectionists. *J Couns Psychol*. 2004;51:192–200. doi:10.1037/0022-0167.51.2.192.
13. Lo A, Abbott MJ. Review of the theoretical, empirical, and clinical status of adaptive and maladaptive perfectionism. *Behav Change*. 2013;30:96–116. doi:10.1017/bec.2013.9.
14. Lloyd S, Schmidt U, Khondoker M, Tchanturia K. Can psychological interventions reduce perfectionism? A systematic review and meta-analysis. *Behav Cogn Psychother*. 2015;43:705–731. doi:10.1017/S1352465814000162.
15. Handley AK, Egan SJ, Kane RT, Rees CS. A randomized controlled trial of group cognitive behavioural therapy for perfectionism. *Behav Res Ther*. 2015;68:37–47. doi:10.1016/j.brat.2015.02.006.
16. Shafran R, Wade T, Egan S, et al. Is the devil in the detail? A randomized controlled trial of guided internet-based CBT for perfectionism. *Behav Res Ther*. 2017;95:99–106. doi:10.1016/j.brat.2017.05.014.
17. Ong CW, Lee EB, Krafft J, et al. A randomized controlled trial of acceptance and commitment therapy for clinical perfectionism. *J Obsessive Compuls Relat Disord*. 2019;22:100444. doi:10.1016/j.jocrd.2019.100444.
18. Arana FG, Miracco MC, Galarregui MS, Keegan EG. A brief cognitive behavioural intervention for maladaptive perfectionism in students: a pilot study. *Behav Cogn Psychother*. 2017;45:537–542. doi:10.1017/S1352465817000406.
19. Bento C, Pereira AT, Roque C, Tavares Saraiva JM, Macedo A. Longitudinal effects of an intervention on perfectionism in adolescents. *Psicothema*. 2017;29:317–322. doi:10.7334/psicothema2016.223.
20. Egan SJ, Hine P. Cognitive behavioural treatment of perfectionism: a single-case experimental design series. *Behav Change*. 2008;25:245–258. doi:10.1375/bech.25.4.245.
21. Steele AL, Waite S, Egan SJ, Finnigan J, Handley A, Wade TD. Psychoeducation and group cognitive-behavioural therapy for clinical perfectionism: a case-series evaluation. *Behav Cogn Psychother*. 2013;41:129–143. doi:10.1017/S1352465812000628.
22. Wade TD, Kay E, Valle MK, et al. Internet-based cognitive behaviour therapy for perfectionism: more is better but no need to be prescriptive. *Clin Psychol*. 2019;23:196–205. doi:10.1111/cp.12193.
23. Furchtlehner LM, Schuster R, Laireiter AR. A comparative study of the efficacy of group positive psychotherapy and group cognitive behavioral therapy in the treatment of depressive disorders: a randomized controlled trial. *J Posit Psychol*. 2020;15:832–845. doi:10.1080/17439760.2019.1663250.
24. Matos M, Steindl SR. "You are already all you need to be": a case illustration of compassion-focused therapy for shame and perfectionism. *J Clin Psychol*. 2020;76:2079–2096. doi:10.1002/jclp.23055.
25. Okumura Y, Ichikura K. Efficacy and acceptability of group cognitive behavioral therapy for depression: a systematic review and meta-analysis. *J Affect Disord*. 2014;164:155–164. doi:10.1016/j.jad.2014.04.023.
26. Mitchell JH, Newall C, Broeren S, Hudson JL. The role of perfectionism in cognitive behaviour therapy outcomes for clinically anxious children. *Behav Res Ther*. 2013;51:547–554. doi:10.1016/j.brat.2013.05.015.
27. Slade PD, Coppel DB, Townes BD. Neurocognitive correlates of positive and negative perfectionism. *Int J Neurosci*. 2009;119:1741–1754. doi:10.1080/00207450902915212.
28. Gäde JC, Schermelleh-Engel K, Klein AG. Disentangling the common variance of perfectionistic strivings and perfectionistic concerns: a bifactor model of perfectionism. *Front Psychol*. 2017;8:160. doi:10.3389/fpsyg.2017.00160.
29. Rhéaume J, Freeston MH, Ladouceur R, et al. Functional and dysfunctional perfectionists: are they different on compulsive-like behaviors? *Behav Res Ther*. 2000;38:119–128. doi:10.1016/S0005-7967(98)00203-4.

30. Schulz KF, Altman DG, Moher D. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332. doi:10.1136/bmj.c332.
31. Vickers AJ, Altman DG. Analysing controlled trials with pretreatment and follow-up measurements. *BMJ*. 2001;323:1123–1124. doi:10.1136/bmj.323.7321.1123.
32. Morris SB. Estimating effect sizes from pretest-posttest-control group designs. *Organ Res Methods*. 2008;11:364–386. doi:10.1177/1094428106291059.
33. Senn S. Change from pretreatment and analysis of covariance revisited. *Stat Med*. 2006;25:4334–4344. doi:10.1002/sim.2682.
34. van Breukelen GJP. ANCOVA versus change from pretreatment in nonrandomized studies: the difference. *Multivariate Behav Res*. 2013;48:895–922. doi:10.1080/00273171.2013.831743.
35. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. *Front Psychol*. 2013;4:863. doi:10.3389/fpsyg.2013.00863.
36. Nakagawa S, Cuthill IC. Effect size, confidence interval and statistical significance: a practical guide for biologists. *Biol Rev*. 2007;82:591–605. doi:10.1111/j.1469-185X.2007.00027.x.
37. Cumming G. The new statistics: why and how. *Psychol Sci*. 2014;25:7–29. doi:10.1177/0956797613504966.
38. Wasserstein RL, Lazar NA. The ASA's statement on p-values: context, process, and purpose. *Am Stat*. 2016;70:129–133. doi:10.1080/00031305.2016.1154108.
39. McShane BB, Gal D, Gelman A, Robert C, Tackett JL. Abandon statistical significance. *Am Stat*. 2019;73(sup1):235–245. doi:10.1080/00031305.2018.1527253.
40. White IR, Horton NJ, Carpenter J, Pocock SJ. Strategy for intention-to-treat analysis in randomised trials with missing outcome observations. *BMJ*. 2011;342:d40. doi:10.1136/bmj.d40.
41. Little RJA, D'Agostino R, Cohen ML, et al. The prevention and treatment of missing observations in clinical trials. *N Engl J Med*. 2012;367:1355–1360. doi:10.1056/NEJMsr1203730.
42. Sterne JAC, White IR, Carlin JB, et al. Multiple imputation for missing observations in epidemiological and clinical research: potential and pitfalls. *BMJ*. 2009;338:b2393. doi:10.1136/bmj.b2393.
43. Kraemer HC, Wilson GT, Fairburn CG, Agras WS. Mediators and moderators of treatment effects in randomized clinical trials. *Arch Gen Psychiatry*. 2002;59:877–883. doi:10.1001/archpsyc.59.10.877.