Research Square

Preprints are preliminary reports that have not undergone peer review.
They should not be considered conclusive, used to inform clinical practice,
or referenced by the media as validated information.

Getting Stuck Differently: The Measurement Structure and Within-Person Temporal Dynamics of Repetitive Negative Thinking in Youth

Shiying Li
Central China Normal University
Kaiwen Ren
Jiangsu Normal University
Xiaoyun Liu
3233700799@qq.com

Hunan First Normal University

Research Article

Keywords: repetitive negative thinking, worry, rumination, internalizing symptoms, psychometric structure, within-person temporal dynamics

Posted Date: November 16th, 2026

DOI: https://doi.org/10.21203/rs.3.rs-11282800/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License.
Read Full License

Additional Declarations: No competing interests reported.

Abstract

Repetitive negative thinking (RNT) is widely regarded as a transdiagnostic process in youth internalizing problems, yet shared process characteristics do not necessarily imply a stable common measurement structure or uniform temporal relations with symptoms. We examined whether commonality across these evidential levels is mutually supportive using two independent open datasets. Study 1 reanalyzed item-level data from Dutch adolescents (N = 3,869), including brief and full Penn State Worry Questionnaire items and rumination/reflection items. Ordinal confirmatory factor analysis, exploratory structural equation modeling (ESEM), item response theory/differential item functioning diagnostics, and supplementary network analysis were used to evaluate the robustness of measurement conclusions. Study 2 used three-wave data from the RETHINK prevention trial (N = 365; ages 16–22 years). Free-lag random-intercept cross-lagged panel models examined bidirectional within-person relations between content-independent RNT and transdiagnostic internalizing symptoms, brooding and depression, and worry and anxiety, with adjustment for randomized condition, inverse-probability weighting, and an alternative anxiety measure for robustness checks. In Study 1, brief and full worry scores were highly correlated, but item-level structural conclusions differed. A content-specific three-factor model received relatively strong support, whereas the conventional bifactor model was unstable under the brief-item specification and failed to converge under the full 26-item specification. ESEM indicated substantive cross-loadings and structure related to reverse wording. In Study 2, no stable directional association emerged between content-independent RNT and transdiagnostic internalizing symptoms. Depression predicted later brooding more strongly in the later interval, although this evidence weakened after multiple-comparison correction. Worry–anxiety relations varied depending on whether anxiety was assessed with the GADQ-IV or GAD-7. Overall, shared measurement structure and within-person temporal functioning represent related but non-interchangeable levels of evidence. Commonality at one level should not be generalized to another without direct testing, and stronger theoretical claims about RNT require complete measurement, defensible model specification, and within-person temporal evidence.

1. Introduction

Internalizing problems such as depression and anxiety in adolescence and early adulthood are expressed not only in emotional and behavioral symptoms but are also often accompanied by persistent negative thought processing. Transdiagnostic research has therefore increasingly shifted from asking "which diagnosis does this belong to?" toward asking "which psychological processes maintain distress across different symptom presentations?" Repetitive negative thinking (RNT) emerged as an important construct within this context and refers to persistent, recurrent thinking about negative experiences or anticipated threats that is difficult to disengage from (Ehring & Watkins, 2008; Watkins, 2008). Prospective studies have shown that RNT is associated with subsequent depressive and anxiety symptoms (McEvoy et al., 2019; Spinhoven et al., 2018), giving it both transdiagnostic theoretical relevance and a role as a process variable in risk identification and prevention research.

RNT is not a single form of thought content. Worry is primarily oriented toward future threat and uncertainty, whereas rumination more often concerns distress that has already occurred or is currently being experienced, together with its causes and consequences (Fresco et al., 2002; Nolen-Hoeksema et al., 2008; Newman et al., 2013). Rumination itself can be differentiated into more maladaptive brooding and reflection, the latter involving a more deliberate problem-solving form of self-focus (Treynor et al., 2003). This distinction also indicates that the combination of worry, rumination, and reflection retained in Study 1 is closer to the broad repetitive-thinking structure described in prior literature, whereas Study 2 focuses on more prototypical forms of negative repetitive thinking, including content-independent RNT, brooding, and worry. Prior research has, on the one hand, found substantial shared variance among these constructs and extracted a general repetitive-thinking factor (Topper, Molenaar, et al., 2014; Hur et al., 2017; Castro et al., 2022; Liu et al., 2024). On the other hand, shared variance does not imply that different constructs must form a stable general factor when fully represented at the item level, nor does it imply that they necessarily occupy the same temporal position in symptom development.

Accordingly, the present research does not frame the question as a binary choice between RNT being "completely the same" or "completely different." Instead, it asks whether structural evidence and temporal evidence can substitute for one another. Study 1 restores the full PSWQ within the existing open dataset and original measurement logic and relaxes constraints such as zero cross-loadings to test whether conclusions about shared structure are robust to item selection and model specification. Study 2 retains the prevention context of the original RETHINK trial and uses three-wave data and RI-CLPMs to separate stable between-person differences from within-person change, testing whether broader RNT and specific RNT-symptom pairings show consistent temporal direction. The two studies therefore address, respectively, whether a common structure is robust and whether a common process manifests as a unified within-person temporal relation, providing independent but complementary evidence about the scope of interpretation for transdiagnostic RNT.

2. Literature Review

2.1 Transdiagnostic RNT Framework: Process Commonality and Construct Boundaries

The transdiagnostic RNT framework proposed by Ehring and Watkins (2008) emphasizes that repetitive negative thinking can differ in content across disorders while sharing process characteristics such as repetitiveness, negativity, and subjective difficulty controlling the thought process. At its core, the framework offers a functional account: it seeks to identify a mode of processing that crosses diagnostic boundaries rather than requiring all manifestations of RNT to be psychometrically identical. Process commonality, measurement commonality, and temporal function should therefore be treated as related but distinct theoretical levels; evidence at one level cannot automatically substitute for tests at another.

Within this framework, worry primarily concerns future threats, uncertainty, and possible negative outcomes (Fresco et al., 2002; Newman et al., 2013), whereas rumination involves sustained attention to

current or past distress and its causes, meaning, and consequences (Nolen-Hoeksema et al., 2008). Brooding involves passively comparing one's current state with an unachieved standard, whereas reflection is more consistent with deliberate, self-focused problem solving (Treynor et al., 2003). Content-independent RNT, by contrast, is defined by the mode rather than the topic of thought, and the PTQ operationalizes this process through repetitiveness, intrusiveness, and difficulty disengaging (Ehring et al., 2011). Reflection is therefore not a prototypical form of negative RNT. In Study 1 it is retained as a boundary construct within an existing broad repetitive-thinking model rather than being treated as equivalent to prototypical RNT.

2.2 Measurement Structure: Common Factors and Model Dependence

Structural studies have repeatedly found substantial overlap between worry and rumination and have used bifactor models to represent a general repetitive-thinking factor together with content-specific residual factors (Topper, Molenaar, et al., 2014; Hur et al., 2017). Castro et al. (2022) extended this logic to worry, rumination, and reflection, and Liu et al. (2024) replicated the corresponding structure in a large Dutch adolescent sample. These findings support the presence of shared variance among forms of repetitive thinking, but they are primarily psychometric evidence about how covariance is organized by a model rather than proof that all forms of repetitive thinking are functionally equivalent.

Interpretation of a common structure may also depend on item representation and model constraints. Bifactor models can separate general from domain-specific variance, but when the general factor absorbs wording effects, content overlap, or other heterogeneous covariance, the resulting solution may become unstable or difficult to interpret (Reise, 2012; Reise et al., 2010). The adolescent replication by Liu et al. (2024) used five PSWQ worry items, whereas the public archive actually contains the complete 16-item PSWQ. Restoring the full item pool is therefore not simply a matter of adding items; it tests whether the previously reported common structure depends on a simplified representation of the worry construct.

Addressing this question requires consideration of both restrictive and more flexible structures. ESEM permits theoretically plausible cross-loadings and can therefore test whether the zero-cross-loading assumption of CFA forces variance attributable to wording or content overlap into a general or specific factor (Asparouhov & Muthen, 2009; Marsh et al., 2014). In the present study, IRT, DIF, and network analysis are used as measurement diagnostics to provide supplementary information about item properties, group-related item functioning, and inter-item connections rather than to establish separate theoretical mechanisms. In this way, the technical analyses in Study 1 all address the same question: does the proposed common structure remain robust under more complete and less restrictive measurement conditions?

2.3 Temporal Direction: Vulnerability and Symptom Feedback

RNT is often positioned upstream of internalizing symptoms. Response Styles Theory proposes that rumination prolongs depressed mood by sustaining attention to negative material and interfering with adaptive problem solving (Nolen-Hoeksema, 1991; Nolen-Hoeksema et al., 2008). Worry has likewise been conceptualized as a repetitive verbal process that maintains anticipated threat and reduces adaptive emotional processing (Newman et al., 2013). Prospective research further indicates that RNT can precede subsequent depressive and anxiety symptoms (Hankin, 2008; McEvoy et al., 2019; Raes, 2012; Spinhoven et al., 2018), providing a clear theoretical basis for an RNT -> symptom cognitive-vulnerability pathway.

Temporal relations, however, need not be unidirectional. Depressive and anxiety states may themselves increase negative self-referential processing, difficulty disengaging attention, and habitual repetitive thinking. The impaired disengagement hypothesis emphasizes difficulty shifting attention away from negative information in depression (Koster et al., 2011), the habit-goal framework proposes that rumination repeatedly triggered by internal or contextual cues can become increasingly automatic (Watkins & Nolen-Hoeksema, 2014), and scar perspectives further suggest that psychopathological states may leave relatively enduring cognitive tendencies (Just et al., 2001). Together, these accounts support the possibility of symptom -> RNT pathways or bidirectional feedback, although the present data can test only temporal ordering and cannot directly identify these specific mechanisms.

The key issue in longitudinal research is therefore not simply whether two variables are associated, but whether stable between-person differences can be separated from deviations within individuals over time. RI-CLPMs use random intercepts to separate stable individual differences, allowing cross-lagged paths to more closely test whether a person being above their own typical level at one occasion is related to a subsequent deviation in another construct (Hamaker et al., 2015; Mulder & Hamaker, 2021). This framework permits comparison of RNT preceding symptoms, symptoms preceding RNT, bidirectional relations, or weaker temporal coupling, providing the methodological basis for Study 2's examination of temporal direction.

2.4 The Present Study: Structural Robustness and Within-Person Dynamics

Existing evidence already shows that forms of repetitive thinking such as worry and rumination share variance while also retaining specificity. The issue that requires greater caution is whether different levels of evidence can substitute for one another. Measurement research can show how shared covariance is organized by latent-variable models, but it cannot establish that different forms of RNT have the same within-person temporal function. Longitudinal research can reveal temporal associations between RNT and symptoms, but it cannot, in reverse, prove the existence of a stable, unitary measurement structure. The present research therefore examines structural robustness and within-person temporal direction in parallel, without treating either as substitute evidence for the other.

The data-generating contexts of the two open datasets also support this layered treatment. The archive used in Study 1 comes from the school-based recruitment project of Topper et al. (2014), whose original

work focused on the measurement of worry and rumination and the development of brief instruments. Liu et al. (2024) later used the same archive to test a bifactor structure of repetitive thinking involving worry, rumination, and reflection. The RETHINK data used in Study 2, by contrast, were generated within a prevention framework in which RNT was treated as a modifiable transdiagnostic risk process, and the randomized prevention trial tested whether reducing RNT could improve subsequent depression and anxiety (Funk et al., 2023, 2025). The present research does not reconstruct those original studies. Instead, while retaining their respective theoretical contexts, it asks whether the prior measurement-structure conclusions are robust and whether RNT and symptoms show stable within-person temporal direction.

Study 1 conducts a secondary analysis of the public adolescent item-level data used by Liu et al. (2024), restoring the full PSWQ after reproducing the brief worry, rumination, and reflection models, and using strict CFA, ESEM, and supplementary IRT/DIF and network diagnostics to test the sensitivity of structural conclusions to item representation and model restrictions. Study 2 conducts a secondary analysis of the three-wave RETHINK trial data, using free-lag RI-CLPMs to examine bidirectional within-person relations between content-independent RNT and transdiagnostic internalizing symptoms, brooding and depression, and worry and anxiety, and further testing the influence of randomized condition, selective attrition, and anxiety operationalization on the conclusions.

Accordingly, the present research specifies three substantive research questions and one robustness question:

RQ1: When the existing brief representation of worry is expanded to the full 16-item PSWQ and measurement restrictions such as zero cross-loadings are relaxed, do conclusions about a general repetitive-thinking structure remain robust?

RQ2: After stable between-person differences are separated, is there a stable directional within-person relation between content-independent RNT and transdiagnostic internalizing symptoms, or is their temporal association weak or bidirectional?

RQ3: Across the specific pairings examined here, do brooding-depression and worry-anxiety show the same temporal direction, or do they exhibit different RNT -> symptom, symptom -> RNT, or bidirectional patterns?

RQ4 (robustness): Do the temporal conclusions remain consistent after adjustment for randomized condition, accounting for selective attrition, and use of an alternative operationalization of anxiety?

3. Methods

3.1 Study 1: Data Source, Original Context, and Measures

Study 1 was a secondary analysis of the openly archived adolescent dataset used by Liu et al. (2024), which originated from the school-based recruitment project of Topper, Emmelkamp, et al. (2014). The original project focused on the measurement of worry and rumination and contributed to the development of brief PSWQ/RRS measures and subsequent risk-screening contexts. Liu et al. (2024) later used the same public archive to examine a bifactor structure of repetitive thinking involving worry, rumination, and reflection. The present study does not reconstruct the original study procedure; rather, within this existing measurement context, it tests the robustness of prior structural conclusions to item representation and model specification. The original report described 3,906 adolescents recruited from 13 secondary schools in the Netherlands, 53.2% of whom were female, with a mean age of 15.85 years (SD = 0.89; range = 14–19 years). The public CSV file available for the present reanalysis contained 3,869 rows rather than the 3,906 participants reported in the publication. Because the archive contains no variable that can explain this 37-case discrepancy, all analyses in the present study report the sample size actually available in the public file and treat the published sample description as contextual information about the source cohort. The public file contains biological-sex coding and item-level responses for the PSWQ and RRS, but no individual-level age or other demographic variables. There were 3,673 complete cases for the 15-item set used to reproduce the brief-scale models and 3,476 complete cases for the full 26-item extension.

Worry was assessed with the 16-item PSWQ, which measures excessive and difficult-to-control worry (Meyer et al., 1990). The public archive had already coded all PSWQ items in the scoring direction used by the original analysis; therefore, no additional reverse scoring was applied. We examined both the five-item brief subset used in the original bifactor replication (items 4, 5, 6, 12, and 13) and all 16 PSWQ items. The brief subset was retained to reproduce the original modeling logic, whereas the full item set addressed the substantive question of whether structural conclusions remained robust when worry was represented more fully. Following the source code, RRS items were used to represent five rumination items (9, 12, 13, 24, and 26) and five reflection items (6, 10, 11, 17, and 18). In the current public file, internal consistency was high for the full PSWQ (alpha = .905) and brief PSWQ (alpha = .857), and acceptable for the rumination (alpha = .779) and reflection (alpha = .767) subscales. These estimates closely matched the reliability profile reported for the original cohort by Liu et al. (2024).

3.2 Study 1: Measurement-Structure Analytical Strategy

Study 1 proceeded from restrictive to more flexible measurement models. First, ordinal confirmatory factor analyses were estimated with robust weighted least squares (WLSMV), treating item responses as ordered indicators. For the original 15-item set, we compared a single-factor model, a content-specific three-factor model distinguishing worry, rumination, and reflection, a valence model combining worry and rumination while retaining reflection separately, and a conventional content-specific bifactor model. We then repeated the content-specific and bifactor modeling logic after restoring all 16 PSWQ items, yielding a 26-item model. Model evaluation treated the CFI, TLI, RMSEA, and SRMR as continuous diagnostic indices rather than applying rigid pass/fail criteria (Hu & Bentler, 1999). Because bifactor models can show apparently favorable fit even when identification is weak or parameters are anomalous,

model convergence, standard-error estimation, and admissibility were treated as substantive diagnostic criteria (Reise, 2012; Reise et al., 2010).

Second, the full 26-item pool was analyzed using a three-factor ESEM with geomin rotation. ESEM relaxes the conventional independent-clusters assumption and was used to determine whether small cross-loadings or wording-related covariance that would be constrained under CFA were being forced into general or content-specific factors (Asparouhov & Muthen, 2009; Marsh et al., 2014). Sex-related robustness of the full content-specific CFA was examined sequentially through configural, loading, and threshold invariance. Given the large sample and the high sensitivity of chi-square difference tests, greater emphasis was placed on changes in CFI and RMSEA (Chen, 2007; Cheung & Rensvold, 2002). Supplementary diagnostics included a graded-response IRT model, sex DIF tests with Benjamini-Hochberg correction, and an EBICglasso item network based on ordinal correlations (Chalmers, 2012; Epskamp et al., 2012, 2018). Because the network solution was relatively dense, network results were treated descriptively only.

3.3 Study 2: Trial Context, Sample, and Measures

Study 2 reanalyzed participant-level data from RETHINK. RETHINK was a preregistered three-arm randomized prevention trial based on the premise that RNT may function as a transdiagnostic risk process for depression and anxiety and as a potential prevention target; the original trial therefore used app-based interventions designed to reduce RNT (Funk et al., 2023, 2025). The sample comprised 365 participants aged 16–22 years who had elevated RNT at enrollment but no current depressive or anxiety disorder. Participants were randomized to a waitlist control condition (n = 122), a concreteness-training intervention (n = 121), or a full RNT-focused intervention (n = 122). Both active interventions lasted six weeks, with assessments conducted at baseline, immediately after the six-week intervention, and at the 18-week follow-up, producing unequal observation intervals of approximately six and 12 weeks. The original trial found no overall advantage of either app over the waitlist on the primary RNT, depression, or anxiety outcomes (Funk et al., 2025). The present study does not reinterpret treatment efficacy; instead, the randomized context is treated as part of the data-generating setting and as exogenous adjustment information for examining within-person temporal relations between RNT and symptoms.

The primary RNT variables comprised three conceptually distinct measures. Content-independent RNT was assessed with the 15-item PTQ, which is designed to capture repetitiveness, intrusiveness, and difficulty disengaging without specifying the topic of thought (Ehring et al., 2011). Brooding was represented by the five-item brooding subscale of the RRS, consistent with the distinction between maladaptive brooding and more reflective forms of self-focused thought (Treynor et al., 2003). Worry was assessed with the 16-item PSWQ (Meyer et al., 1990). Internalizing outcomes were selected to align with the original RETHINK trial and to distinguish transdiagnostic from content-specific tests. Depressive symptoms were measured with the Inventory of Depressive Symptomatology (IDS; Rush et al., 1986). Generalized anxiety was measured primarily with the Generalized Anxiety Disorder Questionnaire-IV (GADQ-IV; Newman et al., 2002), while the seven-item GAD-7 was retained as an alternative operationalization of anxiety in sensitivity analyses (Spitzer et al., 2006). Social anxiety was assessed

with the Social Phobia Inventory (SPIN; Connor et al., 2000). The archive also contains the Patient Health Questionnaire-9 (PHQ-9; Kroenke et al., 2001), but it was used only as a baseline predictor in attrition models and not as a primary depression outcome. For the transdiagnostic analysis, IDS, GADQ-IV, and SPIN total scores were standardized using their baseline means and standard deviations, and their mean was computed at each wave when at least two components were available. This composite was intended to capture broad fluctuations in internalizing symptoms rather than define "transdiagnostic" solely in terms of depression and generalized anxiety.

3.4 Study 2: Within-Person Dynamics and Robustness Analyses

The primary longitudinal analyses used RI-CLPMs estimated with robust maximum likelihood (MLR) and full-information maximum likelihood (FIML) for incomplete repeated measures. RI-CLPMs were selected because conventional cross-lagged models combine stable between-person differences with within-person dynamics; separating random intercepts from occasion-specific deviations provides a closer test of whether a construct being higher than an individual's own expected level precedes a later deviation in another construct (Hamaker et al., 2015; Mulder & Hamaker, 2021). Three primary models were specified: PTQ <-> transdiagnostic internalizing, brooding <-> IDS depression, and PSWQ worry <-> GADQ-IV anxiety. Within each model, random intercepts captured time-invariant between-person variance, whereas autoregressive and cross-lagged paths linked within-person components. Because the two observation intervals differed in length, baseline-to-post-intervention and post-intervention-to-follow-up paths were freely estimated rather than constrained to equality. This allowed the data to express interval-specific coupling rather than imposing a stationarity assumption across the six- and 12-week intervals.

Each primary model was first estimated without treatment covariates and was then re-estimated with dummy-coded randomized assignment to the two active interventions entered as exogenous predictors of the post-intervention and follow-up within-person components. The adjusted models were sensitivity analyses rather than causal treatment-effect models. Directional Wald tests compared the two cross-lagged directions within each interval, and Benjamini-Hochberg correction controlled the false discovery rate within prespecified primary and sensitivity-analysis families (Benjamini & Hochberg, 1995). The relation between worry and GAD-7 was treated as a sensitivity test of outcome operationalization. A total-RRS <-> IDS model produced a negative latent variance and was therefore excluded from substantive interpretation.

Missingness required explicit attention because follow-up participation declined substantially. For most Study 2 measures, data were available for 365 participants at baseline, 249 at post-intervention, and 174 at follow-up; the corresponding numbers for the IDS were 361, 249, and 173. Completion differed across randomized groups: at post-intervention, PTQ data were available for 101/122 participants in the waitlist group, 77/121 in the concreteness-training group, and 71/122 in the full-RNT group; at follow-up, the corresponding numbers were 74, 45, and 55. In addition to FIML, we estimated inverse-probability-of-observation sensitivity models. Logistic regressions predicted post-intervention and follow-up

completion from randomized group and baseline PHQ, GAD-7, IDS, PTQ, brooding, and PSWQ scores. Inverse probabilities were bounded to avoid extreme weights and trimmed at the 99th percentile; lag-specific weighted regressions then estimated both temporal directions using HC3 heteroskedasticity-robust standard errors (Seaman & White, 2013). Analyses were conducted in R 4.6.1 using lavaan 0.7-2 (Rosseel, 2012), psych 2.6.5, mirt 1.47 (Chalmers, 2012), bootnet 1.9.1, and sandwich 3.1-3. The original Study 1 recruitment and the RETHINK trial received the relevant ethics approvals and informed consent. RETHINK was approved by the Ethics Committee of the Department of Psychology at LMU Munich (57_Funk_c), with guardian-consent procedures for minors (Funk et al., 2023). The present study used deidentified public data and involved no new contact with participants. Table 1 summarizes how the two datasets contributed complementary evidence to the overall research design.

Table 1
Overview of the Two-Study Design and Principal Analyses

Feature	**Study 1: Measurement structure**	**Study 2: Temporal dynamics**
Public analytic sample	N = 3,869	N = 365
Developmental range	Source cohort age 14–19 years	Age 16–22 years
Design	Cross-sectional item-level secondary analysis	Three-wave randomized prevention-trial secondary analysis
Core repetitive-thinking/RNT measures	PSWQ; RRS rumination and reflection items	PTQ; RRS brooding; PSWQ worry
Mental-health outcomes	Not available in public file	IDS, GADQ-IV, SPIN; GAD-7 sensitivity analysis
Primary purpose	Test robustness of shared structure to item representation and model specification	Test within-person temporal direction of RNT-symptom relations and sensitivity to operationalization
Primary analyses	Ordinal CFA; full-item ESEM	Free-lag RI-CLPM with FIML/MLR
Robustness/diagnostics	Sex invariance; graded-response IRT/DIF; network analysis	Randomized-condition adjustment; IPW + HC3; alternative anxiety measure

Note. PSWQ = Penn State Worry Questionnaire; RRS = Ruminative Responses Scale; PTQ = Perseverative Thinking Questionnaire; IDS = Inventory of Depressive Symptomatology; GADQ-IV = Generalized Anxiety Disorder Questionnaire-IV; SPIN = Social Phobia Inventory; RI-CLPM = random-intercept cross-lagged panel model; IPW = inverse-probability weighting.

4. Results

4.1 Study 1: Replication of Shared Structure and Full-Item Test

RQ1 first tested whether the common structure supported by the existing brief measurement remained robust when worry was represented in full. The four scale scores were related but not redundant. Spearman correlations were .911 between the full and brief PSWQ, .507 between the full PSWQ and rumination, .403 between the full PSWQ and reflection, .493 between brief worry and rumination, .401 between brief worry and reflection, and .577 between rumination and reflection. Thus, brief and full worry scores were highly consistent at the total-score level, but this correspondence could not by itself establish that the shared structure was equally stable at the item level; the subsequent model comparisons tested this directly.

The competing CFA models are summarized in Table 2. For the 15-item set, the content-specific three-factor model clearly outperformed the single-factor and valence models. Its scaled CFI was .980, TLI was .975, RMSEA was .049, and SRMR was .034; by comparison, the single-factor model had CFI = .835 and RMSEA = .136, and the valence model had CFI = .889 and RMSEA = .113. The original 15-item bifactor model reached an optimization solution and had a low SRMR (.027), but the information matrix could not be inverted, so standard errors and robust test statistics were unavailable. This solution should therefore be treated as estimation-unstable rather than as evidence that the bifactor representation was superior. After restoring all 16 PSWQ items, the content-specific three-factor CFA remained acceptable (scaled CFI = .950, TLI = .945, RMSEA = .058, SRMR = .043), whereas the full 26-item bifactor model failed to converge. Study 1 therefore supports the conclusion that shared variance exists but its conventional bifactor representation is not robust to item expansion, rather than the conclusion that a general repetitive-thinking factor does not exist.

Table 2
Study 1 Measurement-Model Fit

Model	CFI	TLI	RMSEA	SRMR	Scaled chi-square	df	Status
Original 15-item single factor	.835	.808	.136	.105	6552.92	90	Converged
Original 15-item content-specific 3 factor	.980	.975	.049	.034	890.29	87	Converged
Original 15-item valence 2 factor	.889	.869	.113	.086	4451.44	89	Converged
Original 15-item bifactor	NA	NA	NA	.027	NA	75	SE/robust statistics unavailable
Full 26-item content-specific 3 factor	.950	.945	.058	.043	4142.40	296	Converged
Full 26-item ESEM, 3 factor	.959	.946	.057	.033	3413.48	250	Converged
Full 26-item bifactor	NA	NA	NA	NA	NA	NA	Did not converge

Note. CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; ESEM = exploratory structural equation modeling. Dashes indicate that interpretable fit indices were unavailable because of estimation problems.

4.2 Study 1: Cross-Loadings, Wording Effects, and Measurement Diagnostics

Compared with the corresponding strict content-specific CFA, the full 26-item ESEM showed improved fit (CFI = .959, RMSEA = .057, SRMR = .033). More importantly, the loading pattern indicated that a strict general-factor-versus-specific-factor division was insufficient to summarize all item covariance. One factor was defined primarily by reverse-worded PSWQ items, including relatively strong standardized loadings for PSWQ items 1, 3, 8, and 10. A second factor was defined more strongly by core worry items, whereas the third was defined primarily by RRS rumination and reflection items and included multiple cross-loadings. The ESEM results are therefore better interpreted as evidence of measurement complexity jointly involving content, cross-loadings, and wording effects, rather than as direct confirmation or rejection of a single general process. To facilitate direct comparison across competing measurement models, Fig. 1 visualizes the principal CFI, RMSEA, and SRMR estimates for the original 15-item and full 26-item specifications.

Note

Dashed reference lines indicate the fit criteria used in this study: CFI = .95, RMSEA = .06, and SRMR = .08. The original 15-item bifactor solution was unstable because standard errors and robust test statistics were unavailable. The full 26-item bifactor model failed to converge. The open diamond represents the SRMR estimate for the unstable 15-item bifactor solution.

Supplementary diagnostics further indicated the need for caution in measurement interpretation. After correction, several PSWQ items showed sex-related DIF, but the analyses did not establish practically meaningful bias in total scores. DIF was therefore treated as diagnostic information rather than as a basis for sex-specific scoring. The sequence of configural, loading, and threshold models likewise did not yield a simple invariance interpretation, because some constrained models showed better fit on selected indices and the likelihood-ratio testing procedure generated a nesting warning. The EBICglasso network was also relatively dense, so bridge indices were retained only as supplementary description and were not given mechanistic interpretation.

4.3 Study 2: Primary Within-Person Temporal Dynamics

All three primary RI-CLPMs converged successfully and passed post-estimation admissibility checks, and the treatment-adjusted versions showed similarly acceptable fit. Because these models had few degrees of freedom, fit indices were used mainly to screen for obvious misfit, whereas substantive interpretation focused on the cross-lagged estimates. Table 3 reports the primary paths, and Fig. 2 visualizes their standardized effect sizes by lag interval and direction.

Table 3
Primary Free-Lag RI-CLPM Cross-Lagged Associations in Study 2

Model	Lag	Direction	Std. beta	p	BH-adjusted p	Treatment-adjusted Std. beta	Adjusted p
PTQ <-> Internalizing	Baseline->Post	PTQ->Internalizing	.163	.235	.514	.162	.248
PTQ <-> Internalizing	Baseline->Post	Internalizing->PTQ	.209	.318	.544	.197	.324
PTQ <-> Internalizing	Post->Follow-up	PTQ->Internalizing	.121	.237	.514	.122	.231
PTQ <-> Internalizing	Post->Follow-up	Internalizing->PTQ	.118	.426	.639	.122	.418
Brooding <-> IDS	Baseline->Post	Brooding->Depression	.065	.683	.819	.083	.582
Brooding <-> IDS	Baseline->Post	Depression->Brooding	.189	.257	.514	.207	.188
Brooding <-> IDS	Post->Follow-up	Brooding->Depression	.012	.896	.896	.024	.800
Brooding <-> IDS	Post->Follow-up	Depression->Brooding	.352	.005	.063	.351	.007
Worry <-> GADQ-IV	Baseline->Post	Worry->Anxiety	.102	.556	.741	.111	.522
Worry <-> GADQ-IV	Baseline->Post	Anxiety->Worry	.049	.769	.839	.040	.812
Worry <-> GADQ-IV	Post->Follow-up	Worry->Anxiety	.247	.054	.323	.247	.058
Worry <-> GADQ-IV	Post->Follow-up	Anxiety->Worry	.244	.139	.514	.250	.129

Note. Std. beta denotes the standardized within-person cross-lagged coefficient. BH = Benjamini-Hochberg false-discovery-rate correction. Treatment-adjusted models included indicators for randomized assignment to the full-RNT intervention and concreteness training as exogenous predictors of post-intervention and follow-up within-person components. The BH-adjusted values shown are from the unadjusted primary-model family; the adjusted-model family yielded the same substantive conclusions.

Note

Estimates are shown separately for the baseline-to-post-intervention and post-intervention-to-follow-up intervals. Positive values indicate that elevations in the predictor relative to an individual's own expected level were associated with higher subsequent deviations in the outcome.

For RQ2, the PTQ-transdiagnostic internalizing model provided no reliable evidence of a directional within-person association. PTQ elevations did not reliably predict subsequent transdiagnostic internalizing symptoms either from baseline to post-intervention (beta = .163, p = .235) or from post-intervention to follow-up (beta = .121, p = .237), and the reverse internalizing -> PTQ paths were likewise nonsignificant (beta = .209 and .118, respectively). Treatment adjustment materially changed neither the effect sizes nor the statistical inference, and directional Wald tests were nonsignificant in both intervals (unadjusted ps = .500 and .842). Thus, the current data did not identify a stable directional within-person relation between content-independent RNT and transdiagnostic internalizing symptoms.

The brooding-depression model showed the clearest directional asymmetry relevant to RQ3. From baseline to post-intervention, neither brooding -> depression (beta = .065, p = .683) nor depression -> brooding (beta = .189, p = .257) reached statistical reliability. From post-intervention to follow-up, depression predicted subsequent brooding (beta = .352, p = .005), whereas brooding -> depression was near zero (beta = .012, p = .896). The direction and effect sizes were similar after adjustment for randomized condition. However, after BH correction across the primary cross-lagged tests, the p value for depression -> brooding was .063 (.083 in the adjusted model). This result should therefore be regarded as suggestive evidence with a consistent direction rather than as an independent effect established after multiplicity correction in the primary RI-CLPM analyses.

Worry and GADQ-IV showed a different but less clear pattern. From baseline to post-intervention, cross-lagged paths in both directions were small and nonsignificant. From post-intervention to follow-up, worry -> anxiety increased to beta = .247 (p = .054), but anxiety -> worry was also positive (beta = .244, p = .139); the directional Wald test was essentially null (p = .986). Treatment adjustment yielded nearly identical coefficients. Thus, when generalized anxiety was operationalized with the GADQ-IV, the data did not support a clear unidirectional worry-vulnerability interpretation.

4.4 Study 2: Attrition and Outcome-Operationalization Sensitivity Analyses

Attrition was substantial and differed across randomized conditions. Availability of the principal repeated measures was approximately 68% at post-intervention and 48% at follow-up, with lower retention in the active-intervention groups than in the waitlist group. Baseline symptom and RNT measures were not consistently associated with follow-up completion. In addition to FIML, we therefore used IPW to test whether the cross-lagged conclusions depended on differential observation.

The IPW+HC3 results are summarized in Table 4. Consistent with the primary models, the weighted analyses found no reliable bidirectional lagged association between PTQ and transdiagnostic internalizing symptoms, although these nonsignificant findings were not interpreted as direct proof of a

zero effect. The weighted results also strengthened the directional depression -> brooding signal: baseline depression predicted post-intervention brooding at the nominal significance level (b = .155, p = .029), and post-intervention depression more strongly predicted follow-up brooding (b = .213, 95% CI [.103, .323], p < .001, BH p = .002). In contrast, brooding -> depression was nonsignificant in both intervals. For worry and GADQ-IV, worry -> anxiety did not remain significant in either interval after multiplicity correction within the primary analysis family.

With the alternative GAD-7 operationalization, the worry pattern became substantially stronger. In the free-lag RI-CLPM sensitivity model, post-intervention worry predicted GAD-7 symptoms at follow-up (unadjusted beta = .318, p = .026; adjusted beta = .314, p = .036), whereas the reverse path was near zero. In the later lag interval, directional differences were significant in both the unadjusted (p = .026) and adjusted (p = .034) models. These RI-CLPM sensitivity coefficients did not remain significant after the broader correction across all sensitivity cross-lagged paths. However, the IPW+HC3 estimates were robust in both intervals: worry predicted GAD-7 from baseline to post-intervention (b = .308, BH p = .003) and from post-intervention to follow-up (b = .440, BH p < .001), whereas GAD-7 did not predict subsequent worry. Thus, the answer to RQ4 reflects measurement sensitivity rather than simple replication: the prospective worry-anxiety association was substantially stronger when anxiety was represented by the GAD-7 rather than the GADQ-IV. Table 4 presents the lag-specific attrition-weighted results most relevant to robustness.

Table 4
IPW + HC3 Lag-Specific Sensitivity Analyses

Model	Lag	Direction	b	95% CI	p	BH-adjusted p
PTQ <-> Internalizing	Baseline->Post	PTQ->Internalizing	− .015	[-.120, .090]	.780	.780
PTQ <-> Internalizing	Baseline->Post	Internalizing->PTQ	.137	[-.065, .339]	.183	.366
PTQ <-> Internalizing	Post->Follow-up	PTQ->Internalizing	.040	[-.068, .148]	.466	.559
PTQ <-> Internalizing	Post->Follow-up	Internalizing->PTQ	.061	[-.100, .223]	.455	.559
Brooding <-> IDS	Baseline->Post	Brooding->Depression	.049	[-.073, .171]	.429	.559
Brooding <-> IDS	Baseline->Post	Depression->Brooding	.155	[.016, .295]	.029	.118
Brooding <-> IDS	Post->Follow-up	Brooding->Depression	.033	[-.123, .188]	.677	.739
Brooding <-> IDS	Post->Follow-up	Depression->Brooding	.213	[.103, .323]	< .001	.002
Worry <-> GADQ-IV	Baseline->Post	Worry->Anxiety	.133	[-.001, .267]	.051	.154
Worry <-> GADQ-IV	Post->Follow-up	Worry->Anxiety	.166	[.017, .315]	.029	.118
Worry <-> GAD-7 sensitivity analysis	Baseline->Post	Worry->Anxiety	.308	[.133, .483]	< .001	.003
Worry <-> GAD-7 sensitivity analysis	Post->Follow-up	Worry->Anxiety	.440	[.234, .646]	< .001	< .001

Note. IPW = inverse-probability weighting; HC3 = heteroskedasticity-consistent standard errors. Models also included autoregressive outcome terms and randomized-condition indicators. The GAD-7 rows belong to the prespecified sensitivity-analysis family rather than the primary GADQ-IV model.

5. Discussion

5.1 Shared Variance Does Not Equal a Stable Common Measurement Structure

Study 1 first shows that shared variance and a stable general structure are not conclusions at the same evidential level. Worry, rumination, and reflection were moderately correlated, and brief and full PSWQ

scores were also highly consistent, in line with the common components reported in prior repetitive-thinking research. However, when worry was expanded from five selected items to the full 16-item scale and cross-loadings were subsequently allowed, the conventional bifactor representation became estimation-unstable or failed to converge, whereas the content-specific three-factor structure remained estimable. The present findings therefore do not deny the existence of shared variance among forms of repetitive thinking; rather, they show that whether shared variance can be represented as a stable general factor depends on item representation and model specification.

The importance of this result lies in limiting overinterpretation of latent-variable structure. A general factor in a bifactor model is one model-based organization of shared covariance, and it may simultaneously absorb variance contributed by the target process, content overlap, cross-loadings, and wording patterns (Reise, 2012; Reise et al., 2010). In the present ESEM, reverse-worded PSWQ items formed a relatively distinct structure and RRS items showed cross-loadings, indicating that what is represented by a "common factor" is not completely independent of measurement format. Study 1 is therefore better understood as a robustness test of prior structural conclusions than as a verdict on whether RNT itself is unitary.

In addition, the reflection construct retained in Study 1 has a more deliberate, problem-solving character and is not a prototypical form of negative RNT. Retaining it in the model helps reproduce the broad repetitive-thinking structure examined by Liu et al. (2024), but it also underscores an important boundary: the measurement conclusions of Study 1 apply first to the organization of repetitive thinking and cannot be unconditionally equated with a unified latent structure for all forms of RNT. This construct boundary makes integration of the two studies more precise—Study 1 examines the measurement structure of broader repetitive thinking, whereas Study 2 directly examines the temporal relations of prototypical forms of RNT with internalizing symptoms.

5.2 Temporal Direction Is Not a Uniform RNT -> Symptom Pathway

Study 2 shows that no single, stable "RNT precedes symptoms" pattern emerged across the three sets of relations examined here. Content-independent PTQ and transdiagnostic internalizing symptoms showed no reliable directional association in either time interval; brooding-depression showed a more evident depression -> brooding signal in the later interval; and the direction of worry-anxiety relations changed with the operationalization of anxiety. Importantly, these models examine different RNT-symptom pairings. The findings therefore show that the temporal patterns of the pairings studied here are not uniform, but they do not warrant the stronger claim that all RNT types inherently have different temporal mechanisms.

The depression -> brooding direction is compatible with feedback possibilities within Response Styles Theory, the impaired disengagement hypothesis, the habit-goal framework, and scar perspectives (Nolen-Hoeksema et al., 2008; Koster et al., 2011; Watkins & Nolen-Hoeksema, 2014; Just et al., 2001). These theories provide possible explanatory pathways: symptom states may increase the accessibility

of negative material, reduce attentional disengagement, and make repetitive responses easier to trigger again. However, the present study did not directly measure attentional disengagement, habit formation, or cognitive flexibility, so the observed temporal ordering cannot be treated as validation of these specific mechanisms.

On this basis, we retain "symptom-linked cognitive entrenchment" as a descriptive label rather than treating it as a new theory or an established mechanism. The term simply summarizes an observed pattern: after symptoms increase, subsequent repetitive negative thinking may become stronger. Its theoretical value lies in cautioning against positioning RNT invariably upstream of symptoms, while its mechanistic meaning requires further testing through denser longitudinal measurement, behavioral tasks, or experimental manipulation.

Likewise, the two lag intervals should not be interpreted simply as different "optimal timescales." The baseline-to-post-intervention interval was approximately six weeks and coincided with the active intervention period, whereas the post-intervention-to-follow-up interval was approximately 12 weeks and corresponded to the subsequent observation period; attrition was also greater later in the study. Differences between intervals therefore confound interval length, intervention phase, and sample retention. The present study consequently discusses temporal direction without interpreting the six- versus 12-week difference as the time required for cognitive entrenchment or as an optimal intervention window.

5.3 Outcome Operationalization Shapes the Interpretation of Temporal Relations

The worry-anxiety findings further show that interpretation of temporal relations is sensitive to outcome operationalization. With GADQ-IV as the primary outcome, evidence for worry -> anxiety was weak and the direction could not be clearly distinguished. When GAD-7 was used as an alternative operationalization, the worry -> subsequent anxiety pattern was more pronounced, particularly in the IPW+HC3 analyses, where it remained significant after multiplicity correction. The most defensible conclusion is not that one measure reveals the "true mechanism," but that the same theoretical relation changes with how anxiety is measured.

Although both GADQ-IV and GAD-7 concern generalized anxiety, they differ in measurement purpose and content coverage. GADQ-IV is more closely aligned with diagnostic criteria and features of worry, whereas GAD-7 is used primarily to assess dimensional symptom severity (Newman et al., 2002; Spitzer et al., 2006). The two measures may therefore differ in content overlap with the PSWQ, sensitivity to change, and the amount of non-worry symptom variance they contain. The present data are insufficient to determine which specific measurement property produced the difference, but they are sufficient to show that outcome definition is not an incidental detail outside theory testing.

This also suggests that "generality" in transdiagnostic research should be based on convergent evidence across multiple reasonable operationalizations rather than on significance in a single scale or model.

Study 1 shows that item representation can change the interpretation of measurement structure, while Study 2 shows that outcome operationalization can change some temporal associations. Together, the two results support a methodological principle: when a theoretical conclusion is sensitive to measurement choice, that sensitivity is itself a result that should be reported and interpreted.

5.4 Structural and Temporal Evidence Require Level-Specific Interpretation

Across the two studies, the most important integration is not another demonstration that RNT has both commonality and specificity, but the recognition that different levels of evidence answer different questions. A common psychometric structure addresses how different items and constructs share variance, whereas within-person temporal analysis addresses whether a change in one construct relative to an individual's own usual level precedes a change in another construct. These levels are related but not interchangeable: a common factor cannot automatically establish a common temporal function, and a longitudinal path cannot, in reverse, establish a stable common latent structure.

The title phrase "Getting Stuck Differently" can therefore be understood more precisely as follows: similar repetitive negative thinking processes may be linked to internalizing symptoms through different measurement representations and different temporal positions. For some individuals and periods, RNT may function as an upstream risk-related process; in other pairings, symptom states may precede stronger repetitive thinking. The present study does not infer from this that multiple causal mechanisms have been established. Rather, it emphasizes that theoretical interpretations of RNT must specify whether they concern measurement structure, within-person temporal direction, or the operationalization of particular symptoms.

The contribution to the transdiagnostic RNT framework is therefore a calibration across levels of evidence: retaining the explanatory value of RNT as a shared functional process while requiring stronger claims at the structural and temporal levels to receive direct support at their corresponding levels. This approach avoids reducing RNT to a single score or a single pathway, while also avoiding rejection of its transdiagnostic relevance merely because structural or temporal differences are observed. For secondary analyses of open data, this calibration also defines the value of reanalysis—not to redo the original studies, but to test how far existing interpretations remain robust under more complete measurement and more rigorous within-person models.

6. Implications

6.1 Theoretical Implications: Matching Theoretical Claims to Levels of Evidence

First, different claims within RNT theory require evidence at the corresponding level. Functional definitions explain why worry, rumination, and related processes can be placed within a common

transdiagnostic framework; measurement models test how shared variance is organized; and within-person longitudinal models test temporal ordering. When future research advances stronger claims about a "common factor," a "common risk process," or a "unified temporal mechanism," it should specify the direct evidence on which the claim rests rather than allowing results at one level to substitute for tests at another.

Second, temporal direction should be treated as a substantive issue in RNT theory rather than as a by-product of statistical modeling. Existing theories often emphasize RNT as a vulnerability or maintenance process, but the present within-person results remind us that, at least for some RNT-symptom pairings, symptoms may also occupy an upstream temporal position. Future theory should more explicitly distinguish the conditions under which RNT functions as an antecedent, a maintenance process, an outcome, or a bidirectional partner rather than presupposing a single direction.

Third, measurement choice is itself part of theory testing. Item representation in Study 1 and anxiety operationalization in Study 2 each altered some empirical conclusions. The robustness of transdiagnostic theory therefore requires triangulation across reasonable alternative measures and model specifications. Measurement sensitivity should not be dismissed as noise; it should be used to define the scope within which theoretical conclusions are supported.

6.2 Practical Implications: Avoid Equating General RNT Scores with Specific Mechanisms

For school mental-health screening, prevention programs, and clinical assessment, the present findings first suggest a measured interpretation of general RNT scores. A higher RNT score may indicate a tendency toward persistent negative thinking, but that total score alone cannot establish that a particular symptom is being driven by the same mechanism, nor can it determine that RNT is necessarily upstream of symptoms in a given individual. In practice, general RNT indicators, specific worry or brooding, and current symptom status can be recorded as parallel sources of information; however, the present study did not directly test whether such combined assessment is superior to a single indicator.

For intervention design, the present secondary analyses did not compare "early unified intervention" with "later specific intervention," nor did they re-establish the efficacy of the RETHINK app. They therefore cannot be used to recommend a specific treatment strategy. A more cautious practical implication is that when RNT is used as a prevention target, specific forms of RNT, current symptom status, and their changes over time can be recorded together and used to generate subsequent hypotheses. The present study itself cannot establish that this more differentiated assessment improves intervention outcomes.

At the measurement level, brief instruments and broad scales offer practical advantages, but convenience should not substitute for structural and validity testing. Study 1 shows that brief and full PSWQ scores can be highly correlated while item-level structural conclusions still change; Study 2 shows that the temporal relation between the same worry variable and anxiety can differ across anxiety operationalizations. In consequential research or decision contexts, robustness across brief and full

measures and across primary and alternative outcomes should therefore be reported rather than retaining only the most favorable model result.

7. Limitations and Future Research

The present study is first limited by the inherited designs of the secondary open datasets. The public file for Study 1 contains 3,869 observations, whereas the original article reported 3,906 adolescents, and the archive cannot explain the 37-case discrepancy. Individual-level age and broader demographic information are also unavailable. Study 2 spans ages 16–22 years, crossing late adolescence and early adulthood, so developmental interpretations referring to "adolescents" should remain cautious throughout the manuscript. The construct sets also differ across the two studies: Study 1 includes reflection within a broader repetitive-thinking framework, whereas Study 2 focuses on content-independent RNT, brooding, and worry. The two studies therefore provide conceptually complementary evidence rather than a direct replication using the same set of constructs.

Second, all core constructs were measured by self-report. Common-method variance may strengthen associations among variables, and the reverse-wording structure observed in Study 1 also illustrates that item form can affect factor solutions. In Study 2, the transdiagnostic internalizing composite was calculated at each wave when at least two components were available, increasing data use but potentially allowing the composition of the composite score to differ across individuals or waves. The absence of a directional association between PTQ and the internalizing composite may therefore reflect genuinely weak coupling, but it may also be influenced by outcome heterogeneity. Future research could test this with a fixed-composition latent internalizing factor, multimethod assessment, or denser ecological momentary measurement.

Third, RI-CLPMs can separate stable between-person differences from within-person fluctuation, but they cannot establish causal effects. The two lag intervals in RETHINK differ in length and correspond, respectively, to the active intervention period and the subsequent follow-up period, so interval length cannot be fully separated from intervention phase. Follow-up availability fell to approximately 48%, and retention also differed across randomized conditions. FIML and IPW can reduce some missing-data bias, but they cannot eliminate missingness that depends on unobserved variables. In addition, the lag-specific IPW+HC3 regressions are supplementary attrition-sensitivity analyses and are not parameterized identically to the full RI-CLPMs. Agreement between them should therefore be interpreted as directional support rather than as strict replication of the same estimand. Future research requires more waves, more evenly spaced intervals, and preregistered hypotheses about timescale to distinguish genuine dynamic processes from differences introduced by study design.

8. Conclusion

Using two independent open datasets, the present research re-examined the evidential boundaries of transdiagnostic interpretation at two levels: the measurement structure of broad repetitive thinking and

the within-person temporal dynamics of prototypical RNT in relation to internalizing symptoms. Study 1 showed that worry, rumination, and reflection share variance, but conclusions about a general repetitive-thinking structure are sensitive to item representation and model specification; after the full PSWQ was restored, the conventional bifactor representation was no longer stable. Study 2 found no stable directional within-person relation between content-independent RNT and transdiagnostic internalizing symptoms, while identifying a suggestive asymmetric depression-brooding signal and sensitivity of the worry-anxiety relation to outcome operationalization. Together, the two studies show that shared variance, shared process, and shared temporal function should not be treated as synonymous propositions. Transdiagnostic RNT remains a useful functional framework, but stronger interpretations of its structure and dynamic mechanisms require direct support at the corresponding level of evidence.

Declarations

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflict of Interest

The authors declare that they have no competing interests.

Authors' Contributions

Shiying Li: Conceptualization, Methodology, Formal analysis, Visualization, Writing – original draft, Writing – review & editing. Kaiwen Ren: Methodology, Formal analysis, Validation, Writing – review & editing. Xiaoyun Liu: Methodology, Validation, Interpretation, Writing – review & editing. Shiying Li, Kaiwen Ren, and Xiaoyun Liu contributed equally to this work and share first authorship. All authors read and approved the final manuscript.

Acknowledgements

The authors have no additional acknowledgements to declare.

Data Availability

The data analyzed in this study were obtained from two previously collected research datasets. Study 1 used de-identified item-level data from a Dutch adolescent sample (N = 3,869), originally collected by Topper et al. (2014) and subsequently analyzed by Liu et al. (2024). The dataset is available through the

Open Science Framework (OSF) at https://osf.io/ewapy/?view_only=f9068b4e00d2469fbff5b8804a8205ee. Study 2 used de-identified three-wave data from the RETHINK randomized prevention trial (N = 365), reported by Funk et al. (2023, 2025). The anonymized dataset and codebook are linked through the German Clinical Trials Register (DRKS00027384) at https://drks.de/search/de/trial/DRKS00027384, with the associated OSF repository at https://osf.io/va72n/. No new participant data were collected for the present secondary analyses.

Ethics Approval

The present study involved secondary analyses of previously collected, de-identified research datasets and did not involve new participant recruitment, direct contact, or intervention. The original data collection procedures were conducted under the ethical approval and informed consent arrangements described in the respective source publications. Study 1 used previously collected adolescent data associated with Topper et al. (2014) and Liu et al. (2024). Study 2 used data from the RETHINK randomized prevention trial, which received ethical approval from the Ethics Committee of the Department of Psychology at Ludwig Maximilian University of Munich, Germany. Informed consent was obtained from participants in the original trial, with additional parental or legal guardian consent obtained for participants under 18 years of age (Funk et al., 2023)

Clinical Trial Registration

Clinical trial number: not applicable to the present study, as it involved secondary analyses of previously collected data and did not constitute a new clinical trial. The original RETHINK randomized controlled trial was registered with the German Clinical Trials Register (DRKS) under registration number DRKS00027384 on 22 February 2022.

References

1. Asparouhov T, Muthén B. Exploratory structural equation modeling. Struct Equation Modeling: Multidisciplinary J. 2009;16(3):397–438. https://doi.org/10.1080/10705510903008204.
2. Benjamini Y, Hochberg Y. Controlling the false-discovery-rate: A practical and powerful approach to multiple testing. J Roy Stat Soc: Ser B (Methodol). 1995;57(1):289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x.
3. Brosschot JF, Gerin W, Thayer JF. The perseverative cognition hypothesis: A review of worry, prolonged stress-related physiological activation, and health. J Psychosom Res. 2006;60(2):113–24. https://doi.org/10.1016/j.jpsychores.2005.06.074.
4. Burwell RA, Shirk SR. Subtypes of rumination in adolescence: Associations between brooding, reflection, depressive symptoms, and coping. J Clin Child Adolesc Psychol. 2007;36(1):56–65. https://doi.org/10.1080/15374410709336568.

5. Castro AA, Hur J, Berenbaum H. A bifactor approach to modeling the structure of repetitive thinking: Replication and extension. Pers Indiv Differ. 2022;189:111472. https://doi.org/10.1016/j.paid.2021.111472.
6. Chalmers RP. mirt: A multidimensional item-response theory package for the R environment. J Stat Softw. 2012;48(6):1–29. https://doi.org/10.18637/jss.v048.i06.
7. Chen FF. Sensitivity of goodness of fit indexes to lack of measurement invariance. Struct Equation Modeling: Multidisciplinary J. 2007;14(3):464–504. https://doi.org/10.1080/10705510701301834.
8. Cheung GW, Rensvold RB. Evaluating goodness-of-fit indexes for testing measurement invariance. Struct Equation Modeling: Multidisciplinary J. 2002;9(2):233–55. https://doi.org/10.1207/S15328007SEM0902_5.
9. Connor KM, Davidson JRT, Churchill LE, Sherwood A, Foa E, Weisler RH. Psychometric properties of the Social Phobia Inventory (SPIN): New self-rating scale. Br J Psychiatry. 2000;176(4):379–86. https://doi.org/10.1192/bjp.176.4.379.
10. Drost J, van der Does W, van Hemert AM, Penninx BWJH, Spinhoven P. Repetitive negative thinking as a transdiagnostic factor in depression and anxiety: A conceptual replication. Behav Res Ther. 2014;63:177–83. https://doi.org/10.1016/j.brat.2014.06.004.
11. Ehring T, Watkins ER. Repetitive negative thinking as a transdiagnostic process. Int J Cogn Therapy. 2008;1(3):192–205. https://doi.org/10.1521/ijct.2008.1.3.192.
12. Ehring T, Zetsche U, Weidacker K, Wahl K, Schönfeld S, Ehlers A. The Perseverative Thinking Questionnaire (PTQ): Validation of a content-independent measure of repetitive negative thinking. J Behav Ther Exp Psychiatry. 2011;42(2):225–32. https://doi.org/10.1016/j.jbtep.2010.12.003.
13. Epskamp S, Borsboom D, Fried EI. Estimating psychological networks and their accuracy: A tutorial paper. Behav Res Methods. 2018;50:195–212. https://doi.org/10.3758/s13428-017-0862-1.
14. Epskamp S, Cramer AOJ, Waldorp LJ, Schmittmann VD, Borsboom D. qgraph: Network visualizations of relationships in psychometric data. J Stat Softw. 2012;48(4):1–18. https://doi.org/10.18637/jss.v048.i04.
15. Fresco DM, Frankel AN, Mennin DS, Turk CL, Heimberg RG. Distinct and overlapping features of rumination and worry: The relationship of cognitive production to negative affective states. Cogn Therapy Res. 2002;26:179–88. https://doi.org/10.1023/A:1014517718949.
16. Funk J, Kopf-Beck J, Takano K, Watkins E, Ehring T. Can an app designed to reduce repetitive negative thinking decrease depression and anxiety in young people? Results from a randomized controlled prevention trial. J Behav Ther Exp Psychiatry. 2025;87:102014. https://doi.org/10.1016/j.jbtep.2024.102014.
17. Funk J, Kopf-Beck J, Watkins E, Ehring T. Does an app designed to reduce repetitive negative thinking decrease depression and anxiety in young people? (RETHINK): A randomized controlled prevention trial. Trials. 2023;24:295. https://doi.org/10.1186/s13063-023-07295-z.
18. Funk J, Takano K, Schumm H, Ehring T. The bi-factor model of repetitive negative thinking: Common vs. unique factors as predictors of depression and anxiety. J Behav Ther Exp Psychiatry.

2022;77:101781. https://doi.org/10.1016/j.jbtep.2022.101781.

19. Hamaker EL, Kuiper RM, Grasman RPPP. A critique of the cross-lagged panel model. Psychol Methods. 2015;20(1):102–16. https://doi.org/10.1037/a0038889.
20. Hankin BL. Rumination and depression in adolescence: Investigating symptom specificity in a multiwave prospective study. J Clin Child Adolesc Psychol. 2008;37(4):701–13. https://doi.org/10.1080/15374410802359627.
21. Hilt LM, Leitzke BT, Pollak SD. Cognitive control and rumination in youth: The importance of emotion. J Experimental Psychopathol. 2014;5(3):302–13. https://doi.org/10.5127/jep.038113.
22. Hu L-t, Bentler PM. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Struct Equation Modeling: Multidisciplinary J. 1999;6(1):1–55. https://doi.org/10.1080/10705519909540118.
23. Hur J, Heller W, Kern JL, Berenbaum H. A bi-factor approach to modeling the structure of worry and rumination. J Experimental Psychopathol. 2017;8(3):252–64. https://doi.org/10.5127/jep.057116.
24. Joormann J, Gotlib IH. Emotion regulation in depression: Relation to cognitive inhibition. Cogn Emot. 2010;24(2):281–98. https://doi.org/10.1080/02699930903407948.
25. Just N, Abramson LY, Alloy LB. Remitted depression studies as tests of the cognitive-vulnerability hypotheses of depression onset: A critique and conceptual analysis. Clin Psychol Rev. 2001;21(1):63–83. https://doi.org/10.1016/S0272-7358(99)00035-5.
26. Koscinski B, Accorso C, Albanese B, Schmidt NB, Allan NP. A replication and extension of bifactor modelling of perseverative thought in an at-risk community sample: Exploring sex differences in the structure of PT. Br J Clin Psychol. 2023;62:717–30. https://doi.org/10.1111/bjc.12434.
27. Koster EHW, De Lissnyder E, Derakshan N, De Raedt R. Understanding depressive rumination from a cognitive science perspective: The impaired disengagement hypothesis. Clin Psychol Rev. 2011;31(1):138–45. https://doi.org/10.1016/j.cpr.2010.08.005.
28. Kroenke K, Spitzer RL, Williams JBW. The PHQ-9: Validity of a brief depression severity measure. J Gen Intern Med. 2001;16(9):606–13. https://doi.org/10.1046/j.1525-1497.2001.016009606.x.
29. Liu K, Wright ME, Cole LA, Warren SL. Bi-factor model of repetitive thinking: Replication and extension. Pers Indiv Differ. 2024;226:112692. https://doi.org/10.1016/j.paid.2024.112692.
30. Marsh HW, Morin AJS, Parker PD, Kaur G. Exploratory structural equation modeling: An integration of the best features of exploratory and confirmatory factor analysis. Ann Rev Clin Psychol. 2014;10:85–110. https://doi.org/10.1146/annurev-clinpsy-032813-153700.
31. McEvoy PM, Brans S. Common versus unique variance across measures of worry and rumination: Predictive utility and mediational models for anxiety and depression. Cogn Therapy Res. 2013;37:183–96. https://doi.org/10.1007/s10608-012-9448-5.
32. McEvoy PM, Hyett MP, Ehring T, Johnson SL, Samtani S, Anderson R, Moulds ML. Transdiagnostic assessment of repetitive negative thinking and responses to positive affect: Structure and predictive utility for depression, anxiety, and mania symptoms. J Affect Disord. 2018;232:375–84. https://doi.org/10.1016/j.jad.2018.02.072.

33. McEvoy PM, Mahoney AEJ, Moulds ML. Are worry, rumination, and post-event processing one and the same? Development of the Repetitive Thinking Questionnaire. J Anxiety Disord. 2010;24(5):509–19. https://doi.org/10.1016/j.janxdis.2010.03.008.
34. McEvoy PM, Salmon K, Hyett MP, Jose PE, Gutenbrunner C, Bryson K, Dewhirst M. Repetitive negative thinking as a transdiagnostic predictor of depression and anxiety symptoms in adolescents. Assessment. 2019;26(2):324–35. https://doi.org/10.1177/1073191117693923.
35. McLaughlin KA, Nolen-Hoeksema S. Rumination as a transdiagnostic factor in depression and anxiety. Behav Res Ther. 2011;49(3):186–93. https://doi.org/10.1016/j.brat.2010.12.006.
36. Meyer TJ, Miller ML, Metzger RL, Borkovec TD. Development and validation of the Penn State Worry Questionnaire. Behav Res Ther. 1990;28(6):487–95. https://doi.org/10.1016/0005-7967(90)90135-6.
37. Miranda R, Nolen-Hoeksema S. Brooding and reflection: Rumination predicts suicidal ideation at 1-year follow-up in a community sample. Behav Res Ther. 2007;45(12):3088–95. https://doi.org/10.1016/j.brat.2007.07.015.
38. Moulds ML, McEvoy PM. Repetitive negative thinking as a transdiagnostic cognitive process. Nat Reviews Psychol. 2025;4:127–41. https://doi.org/10.1038/s44159-024-00399-6.
39. Mulder JD, Hamaker EL. Three extensions of the random-intercept cross-lagged panel model. Struct Equation Modeling: Multidisciplinary J. 2021;28(4):638–48. https://doi.org/10.1080/10705511.2020.1784738.
40. Newman MG, Llera SJ, Erickson TM, Przeworski A, Castonguay LG. Worry and generalized anxiety disorder: A review and theoretical synthesis of evidence on nature, etiology, mechanisms, and treatment. Ann Rev Clin Psychol. 2013;9:275–97. https://doi.org/10.1146/annurev-clinpsy-050212-185544.
41. Newman MG, Zuellig AR, Kachin KE, Constantino MJ, Przeworski A, Erickson T, Cashman-McGrath L. Preliminary reliability and validity of the Generalized Anxiety Disorder Questionnaire-IV: A revised self-report diagnostic measure of generalized anxiety disorder. Behav Ther. 2002;33(2):215–33. https://doi.org/10.1016/S0005-7894(02)80026-0.
42. Nolen-Hoeksema S. Responses to depression and their effects on the duration of depressive episodes. J Abnorm Psychol. 1991;100(4):569–82. https://doi.org/10.1037/0021-843X.100.4.569.
43. Nolen-Hoeksema S, Stice E, Wade E, Bohon C. Reciprocal relations between rumination and bulimic, substance abuse, and depressive symptoms in female adolescents. J Abnorm Psychol. 2007;116(1):198–207. https://doi.org/10.1037/0021-843X.116.1.198.
44. Nolen-Hoeksema S, Wisco BE, Lyubomirsky S. Rethinking rumination. Perspect Psychol Sci. 2008;3(5):400–24. https://doi.org/10.1111/j.1745-6924.2008.00088.x.
45. Nolen-Hoeksema S, Watkins ER. A heuristic for developing transdiagnostic models of psychopathology: Explaining multifinality and divergent trajectories. Perspect Psychol Sci. 2011;6(6):589–609. https://doi.org/10.1177/1745691611419672.

46. Raes F. Repetitive negative thinking predicts depressed mood at 3-year follow-up in students. J Psychopathol Behav Assess. 2012;34(4):497–501. https://doi.org/10.1007/s10862-012-9295-4.
47. Ren K, Pang C, Sun J, Xie Z. The (In)congruence between creative family and school environments and Chinese adolescents' creative self-efficacy: The mediating role of growth mindset. Think Skills Creativity. 2026;62:102317. https://doi.org/10.1016/j.tsc.2026.102317.
48. Reise SP. The rediscovery of bifactor measurement models. Multivar Behav Res. 2012;47(5):667–96. https://doi.org/10.1080/00273171.2012.715555.
49. Reise SP, Moore TM, Haviland MG. Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. J Pers Assess. 2010;92(6):544–59. https://doi.org/10.1080/00223891.2010.496477.
50. Roelofs J, Rood L, Meesters C, te Dorsthorst V, Bögels S, Alloy LB, Nolen-Hoeksema S. The influence of rumination and distraction on depressed and anxious mood: A prospective examination of the response styles theory in children and adolescents. Eur Child Adolesc Psychiatry. 2009;18(10):635–42. https://doi.org/10.1007/s00787-009-0026-7.
51. Rood L, Roelofs J, Bögels SM, Nolen-Hoeksema S, Schouten E. The influence of emotion-focused rumination and distraction on depressive symptoms in non-clinical youth: A meta-analytic review. Clin Psychol Rev. 2009;29(7):607–16. https://doi.org/10.1016/j.cpr.2009.07.001.
52. Rosseel Y. lavaan: An R package for structural equation modeling. J Stat Softw. 2012;48(2):1–36. https://doi.org/10.18637/jss.v048.i02.
53. Rush AJ, Giles DE, Schlesser MA, Fulton CL, Weissenburger J, Burns C. The Inventory for Depressive Symptomatology (IDS): Preliminary findings. Psychiatry Res. 1986;18(1):65–87. https://doi.org/10.1016/0165-1781(86)90060-0.
54. Seaman SR, White IR. Review of inverse probability weighting for dealing with missing data. Stat Methods Med Res. 2013;22(3):278–95. https://doi.org/10.1177/0962280210395740.
55. Spinhoven P, Drost J, van Hemert B, Penninx BW. Common rather than unique aspects of repetitive negative thinking are related to depressive and anxiety disorders and symptoms. J Anxiety Disord. 2015;33:45–52. https://doi.org/10.1016/j.janxdis.2015.05.001.
56. Spinhoven P, van Hemert AM, Penninx BW. Repetitive negative thinking as a predictor of depression and anxiety: A longitudinal cohort study. J Affect Disord. 2018;241:216–25. https://doi.org/10.1016/j.jad.2018.08.037.
57. Spinhoven P, van Hemert AM, Penninx BW. Repetitive negative thinking as a mediator in prospective cross-disorder associations between anxiety and depression disorders and their symptoms. J Behav Ther Exp Psychiatry. 2019;63:6–11. https://doi.org/10.1016/j.jbtep.2018.11.007.
58. Spitzer RL, Kroenke K, Williams JBW, Löwe B. A brief measure for assessing generalized anxiety disorder: The GAD-7. Arch Intern Med. 2006;166(10):1092–7. https://doi.org/10.1001/archinte.166.10.1092.
59. Topper M, Emmelkamp PMG, Watkins E, Ehring T. Development and assessment of brief versions of the Penn State Worry Questionnaire and the Ruminative Response Scale. Br J Clin Psychol.

2014;53(4):402–21. https://doi.org/10.1111/bjc.12052.

60. Topper M, Emmelkamp PMG, Watkins E, Ehring T. Prevention of anxiety disorders and depression by targeting excessive worry and rumination in adolescents and young adults: A randomized controlled trial. Behav Res Ther. 2017;90:123–36. https://doi.org/10.1016/j.brat.2016.12.015.
61. Topper M, Molenaar D, Emmelkamp PMG, Ehring T. Are rumination and worry two sides of the same coin? A structural equation modelling approach. J Experimental Psychopathol. 2014;5(3):363–81. https://doi.org/10.5127/jep.038813.
62. Treynor W, Gonzalez R, Nolen-Hoeksema S. Rumination reconsidered: A psychometric analysis. Cogn Therapy Res. 2003;27:247–59. https://doi.org/10.1023/A:1023910315561.
63. van Ettekoven KM, Rasing SPA, Vermulst AA, Engels RCME, Kindt KCM, Creemers DH M. Cross-lagged associations between depressive symptoms and response style in adolescents. Int J Environ Res Public Health. 2020;17(4):1380. https://doi.org/10.3390/ijerph17041380.
64. Watkins ER. Constructive and unconstructive repetitive thought. Psychol Bull. 2008;134(2):163–206. https://doi.org/10.1037/0033-2909.134.2.163.
65. Watkins ER, Nolen-Hoeksema S. A habit-goal framework of depressive rumination. J Abnorm Psychol. 2014;123(1):24–34. https://doi.org/10.1037/a0035540.
66. Whitmer AJ, Gotlib IH. An attentional scope model of rumination. Psychol Bull. 2013;139(5):1036–61. https://doi.org/10.1037/a0030923.

Figures

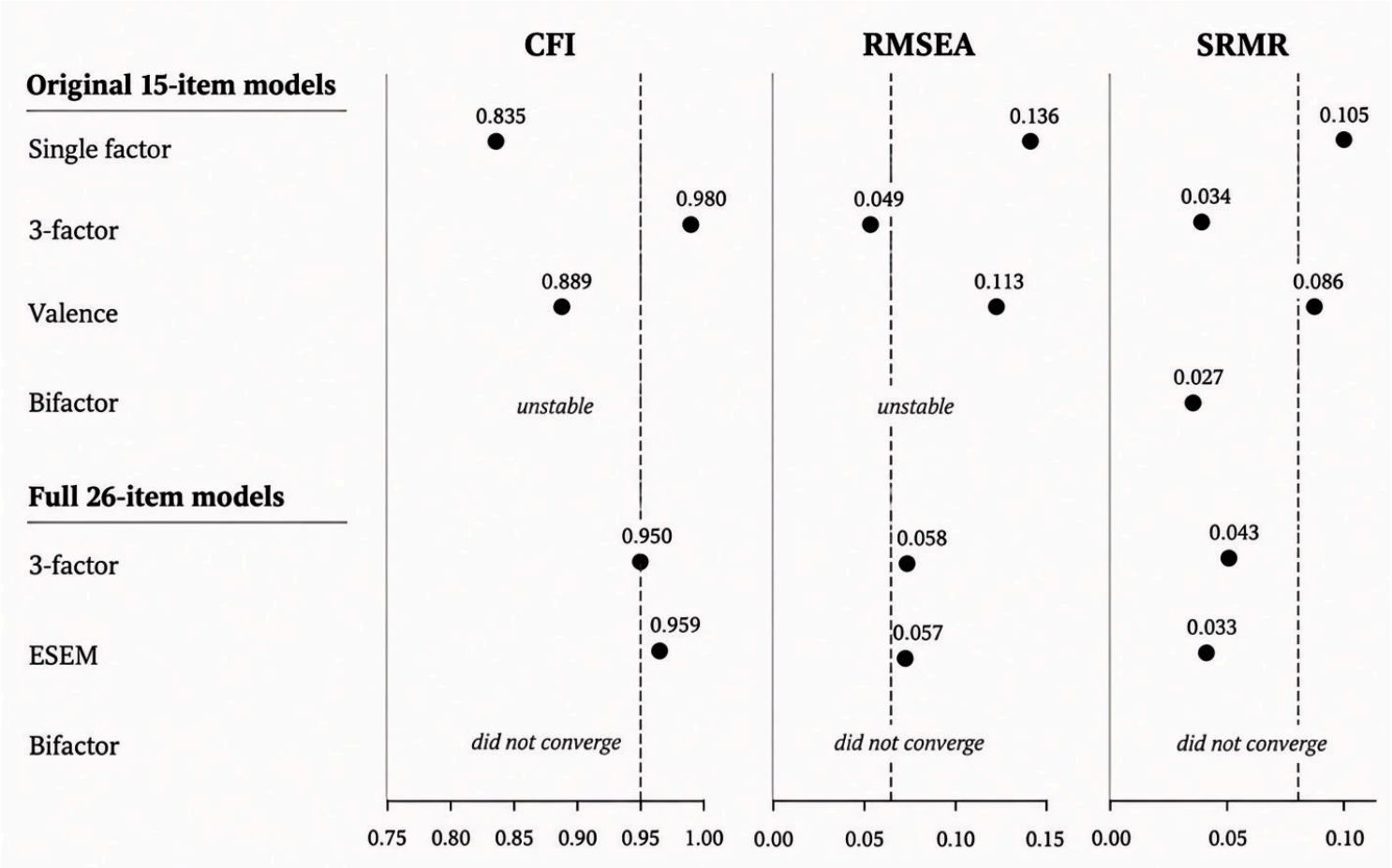


Figure 1

Comparison of Fit Indices Across Study 1 Measurement Models

Note. Dashed reference lines indicate the fit criteria used in this study: CFI = .95, RMSEA = .06, and SRMR = .08. The original 15-item bifactor solution was unstable because standard errors and robust test statistics were unavailable. The full 26-item bifactor model failed to converge. The open diamond represents the SRMR estimate for the unstable 15-item bifactor solution.

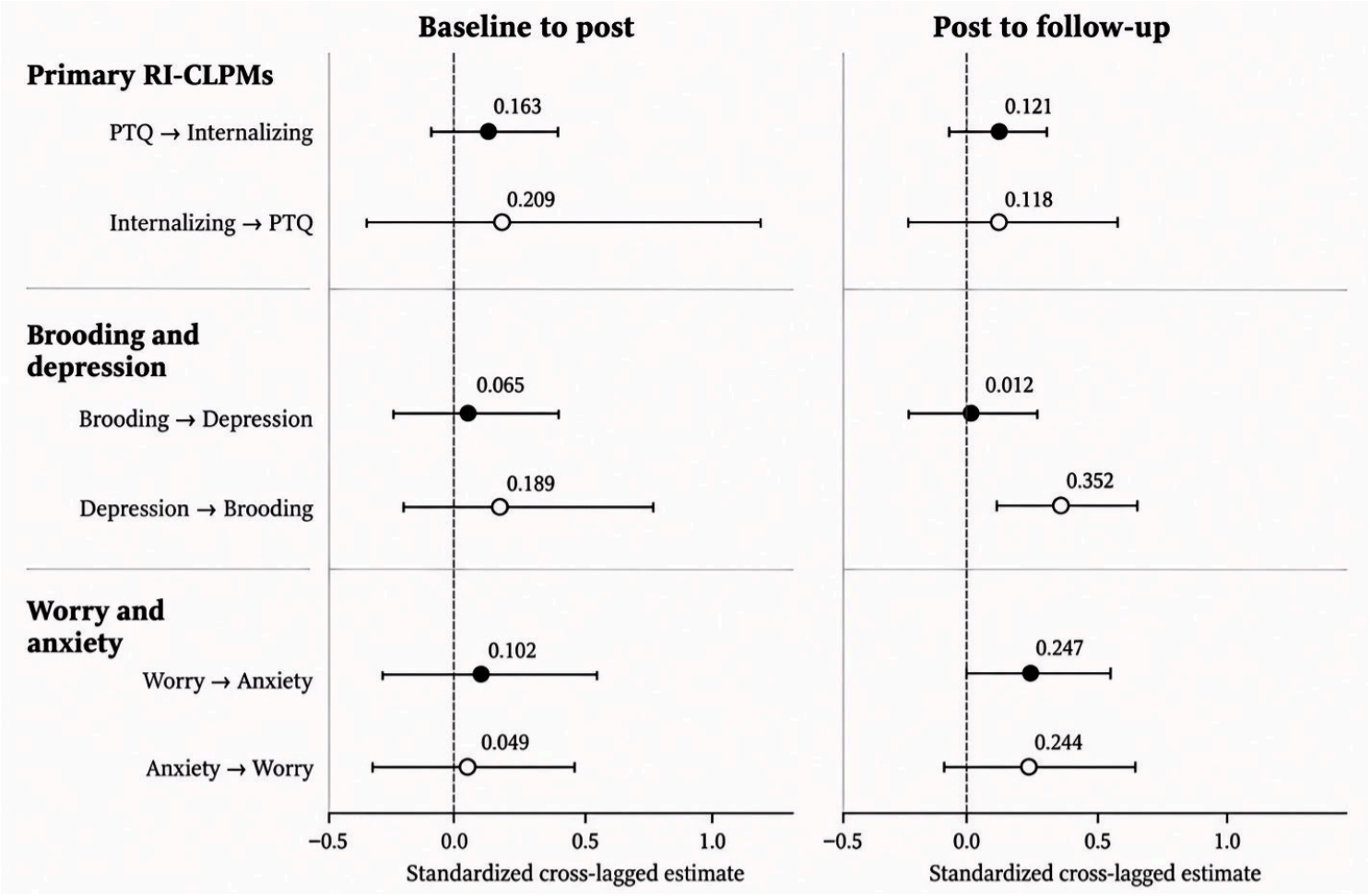


Figure 2

Standardized Within-Person Cross-Lagged Estimates From the Three Primary RI-CLPMs

Note. Estimates are shown separately for the baseline-to-post-intervention and post-intervention-to-follow-up intervals. Positive values indicate that elevations in the predictor relative to an individual's own expected level were associated with higher subsequent deviations in the outcome.