

Development and Limited Validation of a Computerized Adaptive (CAT) Version of the e-QPASS Psychopathology Assessment

Tyler M. Moore [1,2,*], D. K. Adrian Williams [3,5,*], W. Scott Lownsdale [4], Dolores Reidenbach [3], and Shobi S. Ahmed [1,6]

[1]Sama Therapeutics, Cambridge, MA 02141
[2]University of Pennsylvania, Philadelphia, PA 19104
[3]Focused Future LLC, Elizabethtown, PA 17022
[4]QPASSTEST.com, Cañon City, CO 81212
[5]Albert Einstein College of Medicine, Bronx, NY 10461
[6]Sageforce.ai, Newton, MA 02458
[*]These authors contributed equally to this work. Correspondence concerning this article should be addressed to Tyler M. Moore, Sama Therapeutics, 1 Leighton St., Cambridge, MA 02141. E-mail: tyler@sama.ac

Abstract

The e-QPASS (Quick Psychoaffective Symptoms Scan) is a comprehensive 105-item psychopathology assessment that captures three core emotional dimensions underlying mental illness: depression, anxiety, and anger. While clinically valuable, its 10-minute administration time creates barriers for high-throughput clinical and research settings. This study developed a computerized adaptive testing (CAT) version to preserve comprehensive symptom coverage while dramatically reducing respondent burden. We analyzed data from 625 participants (47.8% female, ages 15-79) who completed the full e-QPASS. Exploratory factor analysis identified dimensionality, and Item Response Theory (IRT) Graded Response Models were fitted to calibrate item parameters. CAT simulations determined optimal item administration sequences, with final versions limited to 12 items for general psychopathology and 8 items each for internalizing and externalizing dimensions. Concurrent validity was assessed by examining preservation of known gender differences in externalizing symptoms between full and CAT versions. Factor analysis revealed a robust two-factor structure: Dysregulated Externalizing and Internalizing. IRT models demonstrated acceptable fit (CFI=0.965, RMSEA=0.059). The CAT versions maintained high measurement precision (SE$\leq$0.40) across clinically relevant trait levels while achieving a 10-fold reduction in items. Gender differences in externalizing symptoms were preserved with identical effect sizes across both full and CAT versions, confirming construct validity. The CAT e-QPASS successfully balances comprehensive psychopathology assessment with practical efficiency, offering substantial promise for clinical screening, treatment monitoring, and large-scale research applications where repeated assessments are essential.

Keywords: psychopathology; computerized adaptive testing; psychometrics; gender differences

Research has consistently demonstrated that psychopathology can be organized into a robust two-factor framework consisting of internalizing and externalizing dimensions (Achenbach, 1966; Krueger, 1999). This organizational model emerged from Achenbach's groundbreaking factor-analytic research examining childhood behavioral problems, which repeatedly identified these two broad dimensions across diverse sample populations (Achenbach & Edelbrock, 1978). The internalizing dimension encompasses conditions characterized by internalized distress (such as depression and anxiety disorders), while the externalizing dimension comprises disorders marked by behavioral disinhibition and externally directed expression (including substance use disorders and antisocial behavior) (Krueger & South, 2009). Several theoretical frameworks offer explanations for this bifurcation, including Gray's reinforcement sensitivity theory, which suggests these dimensions reflect distinct neurobiological systems—specifically, the behavioral inhibition system underlying internalizing disorders and the behavioral activation system associated with externalizing psychopathology (Gray, 1987; Bijttebier et al., 2009). Alternative explanatory models emphasize shared genetic vulnerabilities within each spectrum (Kendler et al., 2003) and developmental trajectories involving differential responses to adverse experiences (Cicchetti & Toth, 2009). This parsimonious structure has proven valuable for understanding comorbidity patterns and identifying transdiagnostic mechanisms underlying apparently distinct clinical presentations (Conway et al., 2019).

While traditional approaches conceptualize psychiatric disorders as discrete but overlapping symptom clusters, growing evidence supports the existence of a general psychopathology factor (labeled "p") that spans across diagnostic categories (Caspi et al., 2014; Lahey et al., 2012). The identification of this p factor was foreshadowed by the extensive comorbidity consistently observed among mental disorders (Kessler et al., 2005). Indeed, most factor analyses of psychopathology data reveal that higher-order classifications (such as internalizing and externalizing) are themselves correlated. Multiple researchers, including Krueger (1999) and Krueger et al. (1998), have demonstrated at least moderate correlations among modeled psychopathology factors.

Here, we describe the development of an improved version of an existing psychopathology measure, the e-QPASS (Quick Psychopathology Assessment of Symptoms Scan), which was developed (Lownsdale, 2001) to capture the core emotional dimensions underlying psychopathology. Whereas many existing measures focus on only one domain—often depression, anxiety, or anger—the e-QPASS was designed to integrate all three into a unified framework. This design reflects the theoretical framework in which sadness, fear, and anger often co-occur and contribute synergistically to psychological dysfunction (U.S. National Advisory Mental Health Council, 1995; Russell & Carroll, 1999). Grounded in the Three-Dimensional Affect Model of Psychopathology (Lownsdale, 2001), the e-QPASS is designed to detect elevated levels of these emotions because, when prolonged and intense, they can give rise to maladaptive regulation strategies (e.g., avoidance, compulsive behaviors, substance use) that reflect and reinforce psychiatric illness. The e-QPASS balances diagnostic, treatment, and administrative goals by capturing both emotional distress and associated maladaptive behaviors, providing clinicians with a relatively complete picture of a patient's mental state. With 105 items spanning 14 subscales and yielding scores on 23 constructs, the instrument offers broad coverage of psychopathology in adults.

One potential downside of the e-QPASS is the 10-minute, fixed-length format, which poses barriers for widespread use in high-throughput clinical and research settings (e.g. large genetic studies), particularly those requiring repeated assessments. Ten minutes is extremely fast (even today) by clinical standards, but it is non-negligible in research studies where, often, dozens of measures must be completed in 1-2 hours. While the core value of e-QPASS lies in its comprehensiveness, contemporary psychometric methods offer a path to preserve this depth while drastically reducing respondent burden—a solution explored in the present study. This transition required a psychometric framework capable of evaluating individual items rather than full-scale totals. Item Response Theory (IRT) (Embretson and Reise, 2009; Lord, 1980) provides such a framework, allowing us to model item characteristics and tailor item selection to each respondent's symptom severity. IRT is a psychometric approach that focuses on characteristics of individual test or scale items rather than examining a test or scale holistically. Among the most widely utilized IRT models is the 2-parameter model, expressed by the equation:

$$p_i(\theta) = \frac{1}{1 + e^{-a_i(\theta - b_i)}} \quad (1)$$

In this equation, $p_i(\theta)$ represents the probability of endorsement (or correct response in cognitive assessment), a_i denotes the discrimination parameter, b_i signifies the difficulty parameter, and θ indicates the individual's trait level (in this context, overall mental illness, or OMI). The discrimination parameter (ai) determines how precisely an item can locate an individual along a trait continuum, with higher discrimination values always being preferable. The difficulty parameter (b_i) indicates how elevated an individual must be on the latent trait continuum to have a 50% probability of endorsing the item. For instance, a depression item addressing current suicidal plans would likely be considered "difficult," as one would need to score relatively high on the Depression continuum to endorse it. The more readily an item is endorsed at any given trait level, the lower its difficulty value (i.e., the "easier" it is).

A significant advantage of IRT is its item-level focus, which enables assessment of individual item quality using model-generated parameter estimates. Here, "quality" is determined by the information an item provides at any particular trait level, represented by:

$$I(\theta) = a_i^2 p_i(\theta) q_i(\theta) \qquad (2)$$

Where I(θ) represents item information, a_1 is the discrimination parameter, $p_i(\theta)$ denotes endorsement probability, and $q_i(\theta)$ indicates non-endorsement probability. Three key aspects of this equation warrant attention. First, information invariably increases as a_1 increases, making higher discrimination parameters universally beneficial. Second, the information yielded by an item varies according to person-specific factors. For example, with a_1 fixed at 1.0, an individual low on the trait might exhibit $p_i(\theta)$ = 0.1 and $q_i(\theta)$ = 0.9, yielding total information of 0.1×0.9 = 0.09, while someone slightly higher on the trait might show $p_i(\theta)$ = 0.2 and $q_i(\theta)$ = 0.8, producing total information of 0.2×0.8 = 0.16. Thus, all items generate maximum information when administered to individuals with a 50% endorsement probability—that is, where both $p_i(\theta)$ and $q_i(\theta)$ equal 0.5. An important implication is that items provide optimal information when administered to someone whose trait level (θ) matches the item difficulty (b_i).

These parameters establish the foundation for one of IRT's most innovative applications: computerized adaptive testing (CAT) (Weiss and Kingsbury, 1984; Wainer and Dorans, 2000), which has been previously employed to develop abbreviated and adaptive versions of cognitive and clinical assessments (Moore et al., 2018; Moore et al., 2015; Roalf, Moore et al., 2016). In the CAT methodology, following the initial item administration and response, a scoring algorithm estimates the examinee's trait level and, based on this preliminary estimate, selects the most appropriate subsequent item, where "most appropriate" is defined by information yield. After this next item's administration and response, the algorithm incorporates both responses to estimate the examinee's ability level. The process then identifies the next most appropriate item, continuing iteratively. Assessment concludes when predetermined stopping criteria are met—for example, when the standard error (SE) of measurement reaches a specified threshold.

Notably, because IRT scoring depends on item parameters (ai and bi), endorsing an "easy" item influences a person's score differently than endorsing a "difficult" item. Since all items are calibrated within the same model, examinees can be scored on an identical scale even if they respond to completely different item sets. This feature proves especially valuable for longitudinal research, where item repetition can create methodological challenges.

The current study aimed to develop and validate a CAT assessment of general psychopathology ("p"), along with two specific subfactor scores (internalizing and externalizing). To accomplish this objective, we calibrated 105 items from the e-QPASS clinical assessment to establish an item bank for CAT assessment.

Methods

Participants

The sample comprised N = 625 participants (47.8% female) aged 15–79 years (M = 37.3, SD = 14.3). Participants were recruited through Beag.ly (https://beag.ly), an online research and consumer insight platform where individuals voluntarily register to complete surveys, review products, and provide feedback. All data were collected from participants who consented to Beag.ly's Terms of Use and Privacy Policy, which allow for data contributions to be aggregated and used for research purposes. Because the data were obtained from individuals who had independently enrolled in the Bravo/Beag.ly network and provided information voluntarily outside of a clinical or institutional research setting, no institutional review board (IRB) oversight was required. For participants under the age of 18, guardian assent was included as part of the Beag.ly enrollment process.

Measure

The e-QPASS (Quick Psychoaffective Symptoms Scan) is a 105-item self-report instrument designed to assess a broad range of psychiatric symptoms in adults, with an emphasis on negative affect and associated maladaptive behaviors. Each item is rated on a 5-point Likert scale (1 = not at all to 5 = extremely), and the instrument can typically be completed in approximately 10 minutes. The measure includes three primary scales—depression, anxiety, and anger—and five additional brief scales assessing phobic avoidance, psychoticism, obsessive-compulsivity, suicide risk, and violence risk. Together with its 14 subscales and a global psychopathology index, the e-QPASS enables measurement across 23 clinically relevant constructs. The instrument was originally developed to integrate practical diagnostic, treatment, and administrative utility within a single tool, eliminating the need for multiple overlapping assessments. The theoretical basis of e-QPASS is the Three-Dimensional Affect Model of Psychopathology, which conceptualizes psychopathology as arising from the interaction between intense negative emotional states and maladaptive emotional regulation strategies (Kozak & Cuthbert, 2016; Snyder et al., 2023). This comprehensive scope made the e-QPASS an ideal foundation for building an adaptive assessment using item response theory.

Analysis

The first step in creating the adaptive e-QPASS scales was to determine the dimensionality and item configuration (which items belong to which factor). The ratio of first to second eigenvalues for the e-QPASS was 7.49, suggesting a unidimensional IRT model would likely suffice for modeling the full scale (Embretson & Reise, 2009). However, the scree plot (see Supplementary Figure S1) did suggest some multidimensionality based on the location of the "elbow". While this method clearly suggested three factors, we opted for two factors for two reasons. First the third factor, when extracted, comprised only twelve items, which is insufficient for a CAT item bank. Second, the correlations among the three factors averaged 0.52 (minimum = 0.46), suggesting too much redundancy (overlap) in dimensions being measured to justify producing three scores from the e-QPASS. With the dimensionality finalized at two factors, we conducted an exploratory factor analysis using least-squares extraction and oblique (oblimin) rotation using the psych package (Revelle, 2024) in R.

We then estimated three unidimensional IRT models—specifically Graded Response Models (GRMs) (Samejima, 1969) designed for ordinal responses as here using the mirt package (Chalmers, 2012) in R. One model included all items (overall "p" factor), one included only the internalizing items, and one included only the dysregulated externalizing items. Models were assessed for fit and parameter estimates examined. Model fit was assessed using the comparative fit index (CFI; ¿0.90 acceptable), root mean-square error of approximation (RSMEA; ¡0.08 acceptable), and standardized root mean-square residual (SRMR; ¡0.08 acceptable) (Hu & Bentler, 1999).

With these parameter estimates obtained, we used the e-QPASS items as an item bank for CAT simulation using the catR package (Magis & Raiche, 2012), assessing what would happen if the e-QPASS items were administered adaptively. These included, 1) simulations using the real calibration data to obtain scores on each person for the purposes of validity (see below), and 2) fully simulated sessions for N = 200,000 participants

Table 1: Exploratory Item-Factor Analysis Results (N = 625) for the e-QPASS.

Item	Dysregulated Externalizing	Internalizing
105_Kill_or_physically_harm_another_person	0.95	
103_To_kill_someone	0.94	
83_Breaking_or_destroying_things	0.88	
96_Purging_or_laxatives	0.88	
91_Abusing_medication	0.87	
99_Pulling_out_your_hair	0.86	
97_Having_spending_or_gambling_sprees	0.85	
101_Planning_to_kill_yourself	0.85	
97_Throwing_or_kicking_things	0.85	
104_Killing_or_physically_harm_yourself	0.84	
102_To_beat_or_hit_someone	0.83	
81_Hitting_walls	0.83	
80_Slamming_doors	0.78	
40_A_choking_feeling	0.76	
90_Using_marijuana_or_illegal_drugs	0.75	
26_Having_suicidal_thoughts_or_urges	0.73	
65_Desires_for_harm_to_come_to_someone	0.73	
89_3+_alcoholic_drinks_in_6_hours	0.71	
79_Insulting_others	0.71	
43_Chest_pain_or_tightness	0.70	
86_Feeling_rested_after_[little]_sleep	0.69	
98_Bothered_by_a_sexual_compulsion	0.69	
77_Name-calling	0.69	
93_Certain_objects_or_things	0.68	
31_Alarmed_or_in_danger	0.66	
39_Faintness_dizziness_or_unsteadiness	0.65	
41_Shortness_of_breath_problem_breathing	0.65	
37_Tingling_or_numbness_in_body	0.64	
61_Thoughts_of_getting_even	0.63	
54_Detached_from_your_body	0.62	
36_Trembling_or_shaking_when_it_not_cold	0.61	
64_Feelings_of_hate_towards_someone	0.58	
11_Appetite	0.58	
73_Sulking_or_pouting	0.54	
100_Checking_or_repeating_things	0.53	
35_Sweating_or_hot_flashes_when_not_hot	0.53	
38_Panic_attacks	0.53	
42_Heartbeat_pounding_or_racing	0.52	
55_On_the_lookout_for_danger	0.52	
60_Enraged_or_furious	0.51	
66_Losing_control_or_going_crazy	0.49	
92_Certain_places	0.48	
85_Overeating	0.48	
45_Death_or_dying	0.47	

Continued on next page

Table 1 continued from previous page

Item	Dysregulated Externalizing	Internalizing
85_Being_out_of_your_home	0.47	
51_In_a_dream_or_daze	0.46	0.52
44_Disturbing_thoughts_wont_go_away	0.44	0.40
22_Laughing_or_feeling_amused	0.43	
21_Sitting_or_staying_still	0.43	
88_Having_strong_urges_to_talk_and_talk	0.42	
34_Stomach_discomfort_or_nausea	0.42	0.37
72_Secret_critical_thoughts	0.41	0.36
84_Oversleeping	0.37	
17_Interest_in_things_or_other_people	0.37	
76_Criticizing_others	0.37	0.34
18_Desires_for_sex_or_romance	0.34	
16_Pleasure_or_satisfaction_from_things	0.31	
58_Frustrated		0.82
29_Worried		0.73
3_Tired_or_run_down	-0.42	0.72
27_Nervous_or_anxious		0.72
56_Irritated		0.72
71_Withdrawing_into_yourself_when_upset		0.69
47_Disapproved_of		0.69
66_Undervalued		0.66
57_Disappointed		0.65
49_Rejected		0.65
25_Criticizing_or_blaming_yourself		0.64
9_Like_a_failure		0.64
1_Sad		0.63
33_Tight_tense_muscles		0.62
87_Thoughts_are_racing_through_your_mind		0.61
2_Empty		0.60
63_Problems_letting_go_of_hurt		0.59
11_Helpless		0.59
59_Angry_or_mad		0.58
30_Unable_to_relax_or_calm_yourself		0.58
8_Hopeless		0.58
7_Worthless_or_inadequate		0.57
69_Treated_unfairly		0.57
6_Trapped		0.56
50_Abandoned		0.56
67_Misjudged		0.54
68_Disrespected		0.52
4_Lonely		0.52
75_Not_communicating_with_[enemy]		0.52
48_Ridiculed_or_laughed_at		0.49
55_Troubling_event_all_over_again		0.49

Continued on next page

Table 1 continued from previous page

Item	Dysregulated Externalizing	Internalizing
10_Disgusted_with_yourself		0.49
5_Guilty		0.47
32_Thinking_something_terrible_will_happen	0.36	0.47
62_Feelings_of_bitterness_or_resentment	0.37	0.46
52_Alienated_from_others	0.35	0.46
84_Social_situations	0.31	0.41
28_Scared	0.40	0.41
24_Crying	0.35	0.40
70_Insulted	0.38	0.40
23_Doing_things_more_slowly_than_usual		0.38
74_Not_expressing_anger_when_youre_upset	0.30	0.36
19_Concentrating		0.33
12_Energy		0.32
15_Motivation_or_drive		0.31
20_Making_decisions		0.31
78_Using_sarcasm		0.28
13_Sleep		0.23

Note. Loadings with absolute value < 0.30 omitted unless they are the item's primary loading; rotation = oblimin.

drawn from both normal and uniform distributions to ensure complete coverage of the trait range. The purpose of step 2 was to construct scoring trees of all possible item administration sequences for use in subsequent application of the CAT e-QPASS without having to score participants "on-the-fly" when the e-QPASS is administered in subsequent studies or clinics. Hard-coding these CAT scoring-/administration-trees allows much easier use of the scale without being limited by the computational limits of the system on which it is administered—i.e. the system simply looks up the item administration sequence (and SE) in a table as the scale progresses. Based on these simulations, we limited the length of the full-form CAT to twelve items and the internalizing and externalizing CATs to eight items each. If interested in overall psychopathology ("p"), the 12-item combined would be recommended. If interested in sub-scores, the separate eight-item internalizing and externalizing CATs would be recommended, where "p" could be approximated by taking the mean of those two scores. Note any of these forms could be abbreviated further by using a SE-based stopping rule—e.g. stop when the SE reaches 0.40—but we speculate that the current max lengths are sufficiently short that most users will wish to "max out" the CATs at twelve (combined) or eight (sub-score) adaptively administered items.

Finally, with all real participants scored (step 1 above), we assessed the concurrent validity of the e-QPASS by examining known gender differences in the scores—specifically, that males tend to score higher on externalizing symptoms while females tend to score higher on internalizing symptoms. The question was whether the magnitude and significance of these differences, if found, would differ between the Full (105 total items) and CAT (8 + 8 = 16 total items) versions of the e-QPASS. To test this, we used mixed-effects regression with gender, test type (CAT or Full), and Dimension (internalizing or externalizing) as factors, age as a covariate, and random intercepts for participants. The effect of interest was the three-way interaction of gender*dimension*test_type. We also ran stratified models by dimension, examining the test_type*gender interaction in each to confirm that a non-significant three-way interaction was not due solely to sample size.

Results

The two-factor EFA solution is shown in Table 1. The first factor is most strongly indicated by items 105 ("[In the near future, likely to] kill or physically harm another person"; loading = 0.95), 103 ("[Urges in the past month] to kill someone"; loading = 0.94), 83 ("[In the past seven days, have been] breaking or destroying things"; loading = 0.88), 96 ("[During the past month, have been] purging food or using laxatives"; loading = 0.88), 91 ("[During the past month, have been] abusing medications"; loading = 0.87), and 99 ("[During the past month, have been] pulling out hair, eyebrows, or eyelashes"; loading = 0.86). These and most others on the factor indicate a combination of externalizing and dysregulation; we therefore call this factor "Dysregulated Externalizing". The second factor is most strongly indicated by items 58 ("[Over the past seven days, could be described as] frustrated"; loading = 0.82), 29 ("[Over the past seven days, have been bothered by feeling] worried"; loading = 0.73), 3 ("[Over the past seven days, have been bothered by feeling] tired or run down"; loading = 0.72), 27 ("[Over the past seven days, have been bothered by feeling] nervous or anxious"; loading = 0.72), 56 ("[Over the past seven days, could be described as] irritated"; loading = 0.72), and 71 ("[Over the past seven days, have been] withdrawing into [self] when upset"; loading = 0.69). These and most others on the factor describe internalizing symptoms.

Supplementary Table 1 shows the GRM IRT parameter estimates for the e-QPASS. Fit of the model was acceptable, with a CFI = 0.965, TLI = 0.964, RMSEA = 0.059, and SRMR = 0.073. Discrimination (a) estimates range from very low (0.483; item 3) to very high (5.236; item 103), with a mean of 2.10. An item with a = 2.10 administered to someone exactly at his/her estimated trait level (the goal of CAT) would yield information of 2.102×0.50×0.50 = 1.10. This corresponds to a standard error of 1/sqrt(1.10) = 0.95. Because IRT trait (theta) parameters are in a z metric, that can be interpreted as estimation within 0.95 standard

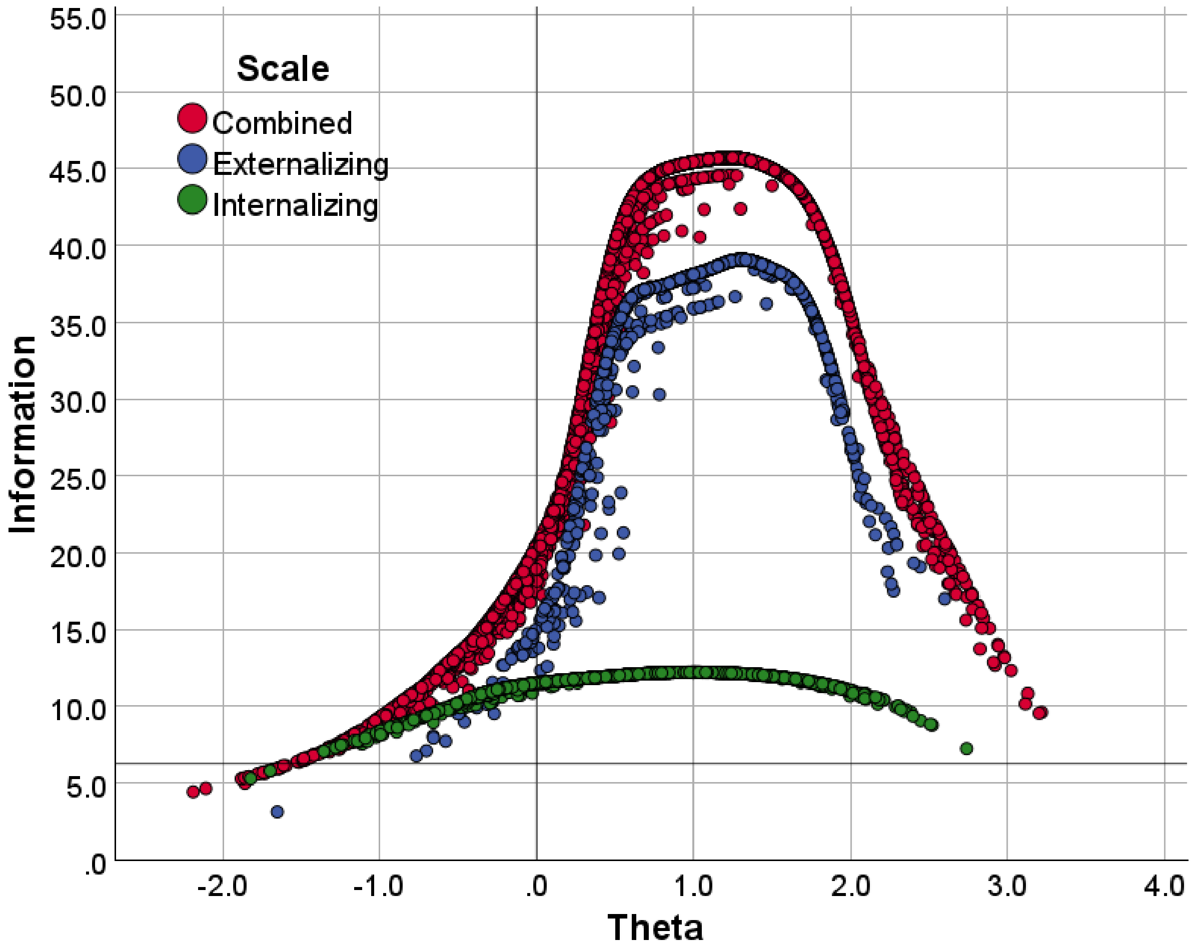

Figure 1: Standard Error-Based Information Functions for 12-item Combined, 8-Item Externalizing, and 8-Item Internalizing Adaptive (CAT) Scales of e-QPASS.

deviations (SDs). GRM threshold (b) parameters range from very low (-4.384; item 3) to very high (4.708; item 12) with a mean of 1.08. This indicates (roughly) that the e-QPASS provides the most information (lowest SE) for participants about one SD above the mean on overall psychopathology.

Figure 1 shows the scale information function for the three e-QPASS scales, where higher values indicate more information (lower SE). It is clear from this function that all three scales produce the most information between 0.50 and 1.50 SDs above the mean on each trait (x-axis, Theta). A horizontal line has been drawn at 6 on the y-axis, roughly corresponding to a standard error of 0.40. This is sometimes used as a maximum acceptable SE, where, when the SD of the trait level is 1.0 (as it is here), the marginal reliability is 1 – SE2 = 1 – 0.402 = 0.84, which is good (using the same standards as the alpha reliability coefficient). Thus, while the e-QPASS internalizing CAT yields far less information than the externalizing and full forms, it remains above the minimum acceptable (6) for quite a wide range of the trait level. Indeed, for part of the trait range (0.25 to 2.00), the externalizing and full CATs maintain information ¿25, equivalent to a SE of 0.20 (extremely precise) with as few as eight items (externalizing) or twelve items (combined).

Figure 2 shows the concurrent validation results examining gender differences in all four score types: CAT-based externalizing, CAT-based internalizing, Full-bank externalizing, and Full-bank internalizing. As predicted, males show higher externalizing scores than females, although the sex difference for internalizing is not significant. Importantly, the interaction between gender and score type is not significant—i.e. the gender difference in externalizing is not significantly larger when using the full item bank than when using the 8-item CAT score.

Table 2 shows the results of these mixed-effects

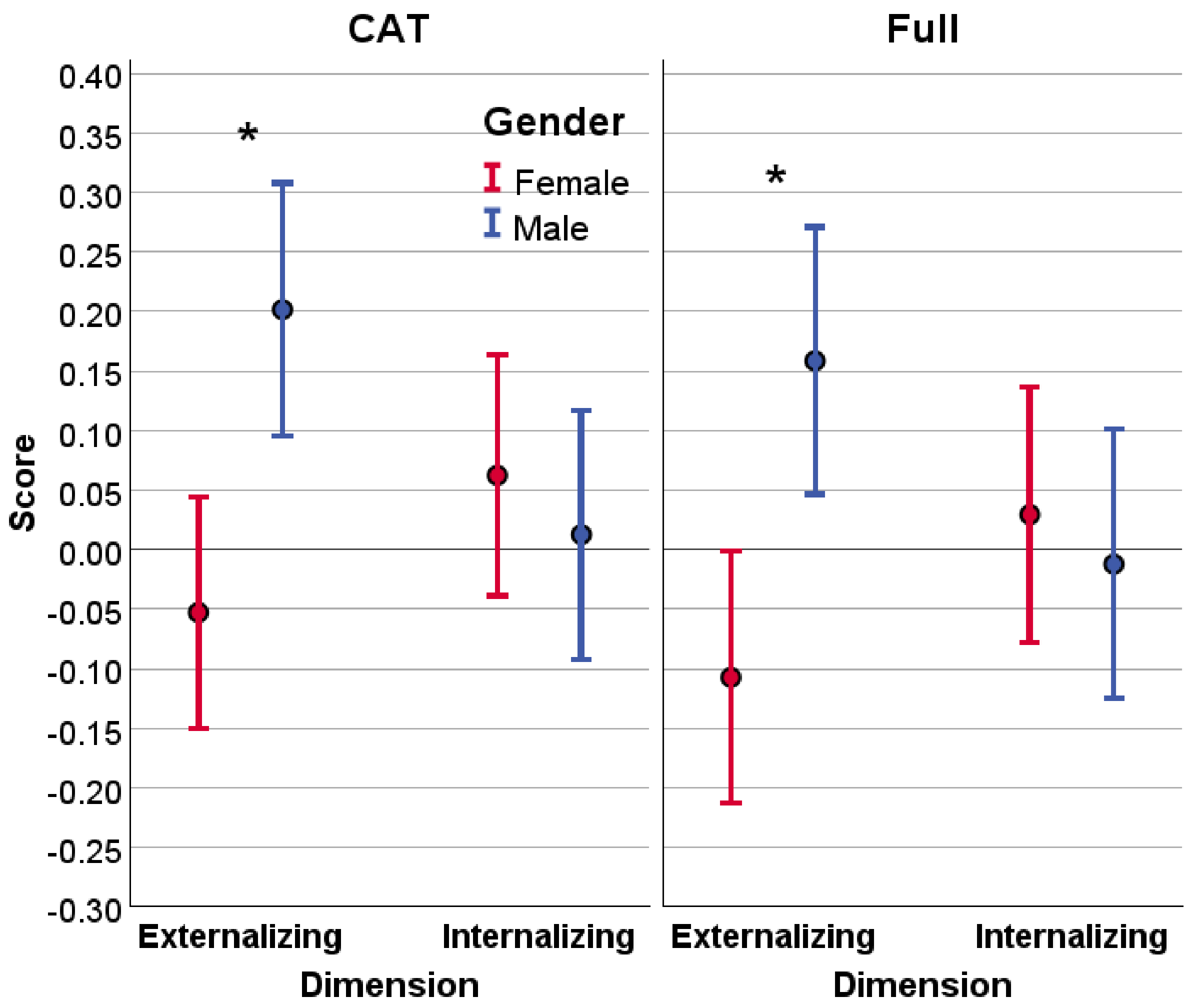

Figure 2: Gender Differences in e-QPASS Scores, by Psychopathology Dimension and Score Type.

regression models revealing a significant main effect of gender for the Externalizing score and no significant interactions between gender and score type (CAT versus full).

Discussion

The successful development of a computerized adaptive version of the e-QPASS represents a significant advancement in psychopathology assessment, offering clinicians and researchers a tool that maintains comprehensive symptom coverage while dramatically reducing administration time. The preservation of psychometric quality with just 12 items (for general psychopathology) or 8 items per dimension represents a 10-fold reduction in respondent burden compared to the original 105-item version. This efficiency gain has profound implications for clinical practice, where time constraints often force clinicians to choose between comprehensive assessment and practical feasibility. The CAT version's ability to provide at least moderately precise measurement (SE $\leq$ 0.40) across a wide range of symptom severity levels makes it particularly valuable for screening in primary care settings, monitoring treatment response, and conducting large-scale epidemiological research where repeated assessments are essential.

Perhaps most theoretically significant is the demonstration that the e-QPASS items naturally organize into a robust two-factor structure that aligns with decades of psychopathology research on internalizing and externalizing dimensions (Achenbach et al., 2016; Achenbach, 1966). The emergence of a ”Dysregulated Externalizing” factor, characterized by items reflecting both behavioral disinhibition and self-regulatory failures, provides empirical support for contemporary models that emphasize emotion dysregulation as a core transdiagnostic mechanism (Aldao et al., 2016). The

Table 2. Mixed-Effects Regression Results Examining Interaction of Gender and Score Type for the e-QPASS.

DV	IV	Estimate	SE	t	p-value
general "p"	Age	-0.018	0.003	-6.6	< 0.0005
	Male_Gender	0.056	0.078	0.7	0.476
	Score_Type	-0.007	0.016	-0.5	0.641
	Male_Gender:Score_Type	0.017	0.022	0.8	0.434
Externalizing	Age	-0.018	0.003	-6.7	< 0.0005
	Male_Gender	0.209	0.078	2.7	0.007
	Score_Type	0.005	0.014	0.4	0.718
	Male_Gender:Score_Type	-0.007	0.020	-0.4	0.725
Internalizing	Age	-0.018	0.003	-6.6	< 0.0005
	Male_Gender	-0.120	0.078	-1.5	0.126
	Score_Type	-0.007	0.017	-0.4	0.686
	Male_Gender:Score_Type	0.012	0.024	0.5	0.626

Note. DV = dependent variable; IV = independent variable; SE = standard error; general "p" = overall psychopathology.

exceptional discrimination parameters achieved by some items (up to 5.236 for homicidal ideation) suggest that certain severe symptoms may serve as particularly powerful indicators of overall psychopathology burden. The scale's ability to maintain high information content (¿25) across trait levels corresponding to 0.25-2.00 standard deviations above the mean indicates that the CAT version is optimally suited for detecting clinically significant symptom levels rather than subclinical variations—a feature that enhances its clinical utility.

The replication of established gender differences in externalizing symptomatology (Lombroso & Ferrero, 1895), with males scoring significantly higher than females, provides important concurrent validation for the CAT version and reinforces the robustness of this well-documented phenomenon (Hicks et al., 2007). The finding that these gender differences were preserved with identical effect sizes across both the full 105-item version and the 16-item CAT version (8 items each for internalizing and externalizing) demonstrates that adaptive testing maintains sensitivity to known group differences while substantially reducing assessment burden. This preservation of construct validity across administration formats is particularly noteworthy given that the CAT version administers different items to different participants based on their response patterns. The absence of significant gender differences in internalizing scores, while inconsistent with some previous research showing female predominance in internalizing disorders, may reflect the specific item composition of the e-QPASS or the particular sample characteristics.

The development of pre-computed scoring trees representing all possible item administration sequences is an innovative methodological contribution that addresses practical implementation barriers commonly associated with CAT systems. By eliminating the need for real-time IRT computations during assessment, this approach makes the CAT version accessible to clinical settings with limited computational resources. This methodological advance removes a significant barrier to widespread CAT adoption and demonstrates how sophisticated psychometric techniques can be made practically accessible to end users without specialized technical expertise.

Some limitations warrant consideration when interpreting these findings. The reliance on a single calibration sample may limit generalizability to populations with different demographic characteristics or symptom presentations. Additionally, the validation evidence presented here is limited to known-groups validity (gender differences) in externalizing symptoms, leaving broader questions about construct validity and criterion-related validity unaddressed. Despite these limitations, however, the CAT version of the e-QPASS represents a significant methodological advance that successfully balances comprehensive psychopathology assessment with practical efficiency, offering substantial promise for both clinical practice and research applications.

Declaration of conflicting interest

T.M.M. and S.S.A. have financial interests, including equity ownership and/or receipt of consulting fees from Sama Therapeutics. D.K.A.W. and D.R. have financial interests, including equity ownership and/or receipt of payment or credit from Focused Future LLC. W.S.L. is owner of QPASSTEST.com. While D.K.A.W. is affiliated with the Albert Einstein College of Medicine, the institution did not participate in, sponsor, nor endorse this work. Likewise, while T.M.M. is affiliated with the University of Pennsylvania and Children's Hospital of Philadelphia, neither institution participated in, sponsored, nor endorses this work.

Funding Statement

This work was supported by Sama Therapeutics.

References

Achenbach, T. M. (1966). The classification of children's psychiatric symptoms: A factor-analytic study. Psychological Monographs: General and Applied, 80(7), 1–37. https://doi.org/10.1037/h0093906

Achenbach, T. M., & Edelbrock, C. S. (1978). The classification of child psychopathology: A review and analysis of empirical efforts. Psychological Bulletin, 85(6), 1275–1301. https://doi.org/10.1037/0033-2909.85.6.1275

Achenbach, T. M., Ivanova, M. Y., Rescorla, L. A., Turner, L. V., & Althoff, R. R. (2016). Internalizing/externalizing problems: Review and recommendations for clinical and research applications. Journal of the American Academy of Child & Adolescent Psychiatry, 55(8), 647–656.

Aldao, A., Gee, D. G., De Los Reyes, A., & Seager, I. (2016). Emotion regulation as a transdiagnostic factor in the development of internalizing and externalizing psychopathology: Current and future directions. Development and Psychopathology, 28(4pt1), 927–946.

Bijttebier, P., Beck, I., Claes, L., & Vandereycken, W. (2009). Gray's Reinforcement Sensitivity Theory as a framework for research on personality–psychopathology associations. Clinical Psychology Review, 29(5), 421–430. https://doi.org/10.1016/j.cpr.2009.04.002

Caspi, A., Houts, R. M., Belsky, D. W., Goldman-Mellor, S., Harrington, H., Israel, S., Meier, M. H., Ramrakha, S., Shalev, I., Poulton, R., & Moffitt, T. E. (2014). The p Factor. Clinical Psychological Science, 2(2), 119–137. https://doi.org/10.1177/2167702613497473

Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software, 48(6), 1–29. https://doi.org/10.18637/jss.v048.i06

Cicchetti, D., & Toth, S. L. (2009). The past achievements and future promises of developmental psychopathology: The coming of age of a discipline. Journal of Child Psychology and Psychiatry, 50(1-2), 16–25. https://doi.org/10.1111/j.1469-7610.2008.01979.x

Conway, C. C., Forbes, M. K., Forbush, K. T., Fried, E. I., Hallquist, M. N., Kotov, R., Mullins-Sweatt, S. N., Shackman, A. J., Skodol, A. E., South, S. C., Sunderland, M., Waszczuk, M. A., Zald, D. H., Afzali, M. H., Bornovalova, M. A., Carragher, N., Docherty, A. R., Jonas, K. G., Krueger, R. F., … Eaton, N. R. (2019). A hierarchical taxonomy of psychopathology can transform mental health research. Perspectives on Psychological Science, 14(3), 419–436. https://doi.org/10.1177/1745691618810696

Embretson, S. E., & Reise, S. P. (2009). Item Response Theory for Psychologists (Reprint ed.). Erlbaum.

Gray, J. A. (1987). The Psychology of Fear and Stress (2nd ed.). Cambridge University Press.

Hicks, B. M., Blonigen, D. M., Kramer, M. D., Krueger, R. F., Patrick, C. J., Iacono, W. G., & McGue, M. (2007). Gender differences and developmental change in externalizing disorders from late adolescence to early adulthood: A longitudinal twin study. Journal of Abnormal Psychology, 116(3), 433.

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling: A Multidisciplinary Journal, 6(1), 1–55.

Kendler, K. S., Prescott, C. A., Myers, J., & Neale, M. C. (2003). The structure of genetic and environmental risk factors for common psychiatric and substance use disorders in men and women. Archives of General Psychiatry, 60(9), 929–937. https://doi.org/10.1001/archpsyc.60.9.929

Kessler, R. C., Chiu, W. T., Demler, O., & Walters, E. E. (2005). Prevalence, severity, and comorbidity of 12-month DSM-IV disorders in the National Comorbidity Survey Replication. Archives of General Psychiatry, 62(6), 617–627. https://doi.org/10.1001/archpsyc.62.6.617

Kozak, M. J., & Cuthbert, B. N. (2016). The NIMH Research Domain Criteria Initiative: Background, issues, and pragmatics. Psychophysiology, 53(3), 286–297.

Krueger, R. F. (1999). The structure of common mental disorders. Archives of General Psychiatry, 56(10), 921–926. https://doi.org/10.1001/archpsyc.56.10.921

Krueger, R. F., Caspi, A., Moffitt, T. E., & Silva, P. A. (1998). The structure and stability of common mental disorders (DSM-III-R): A longitudinal-epidemiological study. Journal of Abnormal Psychology, 107(2), 216–227. https://doi.org/10.1037/0021-843X.107.2.216

Krueger, R. F., & South, S. C. (2009). Externalizing disorders: Cluster 5 of the proposed meta-structure for DSM-V and ICD-11. Psychological Medicine, 39(12), 2061–2070. https://doi.org/10.1017/S0033291709990328

Lahey, B. B., Applegate, B., Hakes, J. K., Zald, D. H., Hariri, A. R., & Rathouz, P. J. (2012). Is there a general factor of prevalent psychopathology during adulthood? Journal of Abnormal Psychology, 121(4), 971–977. https://doi.org/10.1037/a0028355

Lombroso, C., & Ferrero, G. (1895). The Female Offender. D. Appleton.

Lord, F. M. (1980). Applications of Item Response Theory to Practical Testing Problems. Erlbaum.

Lownsdale, W. S. (2001). The Development and Validity of the Psychological Symptoms Scan. Doctoral dissertation, Northern Illinois University.

Magis, D., & Raiche, G. (2012). Random generation of response patterns under computerized adaptive testing with the R package catR. Journal of Statistical Software, 48(8), 1–31. https://doi.org/10.18637/jss.v048.i08

Moore, T. M., Calkins, M. E., Reise, S. P., Gur, R. C., & Gur, R. E. (2018). Development and public release of a computerized adaptive (CAT) version of the Schizotypal Personality Questionnaire. Psychiatry Research, 263, 250–256. https://doi.org/10.1016/j.psychres.2018.02.022

Moore, T. M., Calkins, M. E., Satterthwaite, T. D., Roalf, D. R., Rosen, A. F., Gur, R. C., & Gur, R. E. (2019). Development of a computerized adaptive screening tool for overall psychopathology (“p”). Journal of Psychiatric Research, 116, 26–33. https://doi.org/10.1016/j.jpsychires.2019.05.028

Moore, T. M., Scott, J. C., Reise, S. P., Port, A. M., Jackson, C. T., Ruparel, K., Savitt, A. P., Gur, R. E., & Gur, R. C. (2015). Development of an abbreviated form of the Penn Line Orientation Test using large samples and computerized adaptive test simulation. Psychological Assessment, 27(3), 955–964. https://doi.org/10.1037/pas0000102

Reise, S. P., Cook, K. F., & Moore, T. M. (2015). Evaluating the impact of multidimensionality on unidimensional Item Response Theory model parameters. In S. P. Reise & D. Revicki (Eds.), Handbook of Item Response Theory Modeling (pp. 13–40). Routledge.

Reise, S. P., Moore, T. M., & Maydeu-Olivares, A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional Item Response Theory models. Educational and Psychological Measurement, 71(4), 684–711. https://doi.org/10.1177/0013164410378690

Reise, S. P. (2012). The rediscovery of bifactor measurement models. Multivariate Behavioral Research, 47(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. Journal of Personality Assessment, 92(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Revelle, W. (2024). psych: Procedures for Psychological, Psychometric, and Personality Research (R package version 2.4.2) [Computer software]. Northwestern University. https://CRAN.R-project.org/package=psych

Roalf, D. R., Moore, T. M., Wolk, D. A., Arnold, S. E., Mechanic-Hamilton, D., Rick, J., Kabadi, S., Ruparel, K., Chen-Plotkin, A., Chahine, L. M., Dahodwala, N. A., Duda, J. E., Weintraub, D. A., & Moberg, P. J. (2016). Defining and validating a short form Montreal Cognitive Assessment (s-MoCA) for use in neurodegenerative disease. Journal of Neurology, Neurosurgery & Psychiatry, 87(12), 1303–1310. https://doi.org/10.1136/jnnp-2015-312723

Russell, J. A., & Carroll, J. M. (1999). On the bipolarity of positive and negative affect. Psychological Bulletin, 125(1), 3–30.

Samejima, F. (1969). Estimation of Latent Ability Using a Response Pattern of Graded Scores (Psychometrika Monograph Supplement No. 17). Psychometric Society.

Snyder, H. R., Silton, R. L., Hankin, B. L., Smolker, H. R., Kaiser, R. H., Banich, M. T., … Heller, W. (2023). The dimensional structure of internalizing psychopathology: Relation to diagnostic categories. Clinical Psychological Science, 11(6), 1044–1063.

United States National Advisory Mental Health Council. (1995). Basic Behavioral Science Research for Mental Health: A National Investment (Vol. 95, No. 3682). National Institute of Mental Health.

Wainer, H., & Dorans, N. J. (2000). Computerized Adaptive Testing (2nd ed.). Routledge.

Weiss, D. J., & Kingsbury, G. G. (1984). Application of computerized adaptive testing to educational problems. Journal of Educational Measurement, 21(4), 361–375. https://doi.org/10.1111/j.1745-3984.1984.tb01040.x