Technical Report 1397

An Examination of the Longitudinal Stability of Psychological Measures Contained within the U.S. Army's Global Assessment Tool

Nathaniel J. Ratcliff
Joel Thurston
Joshua R. Goldstein
Vicki A. Lancaster
Stephanie S. Shipp
Sallie Keller
University of Virginia,
Biocomplexity Institute and Initiative,
Social and Decision Analytics Division

Kelly S. Ervin
U.S. Army Research Institute

April 2021

United States Army Research Institute for the Behavioral and Social Sciences

U.S. Army Research Institute
for the Behavioral and Social Sciences

Department of the Army
Deputy Chief of Staff, G1

Authorized and approved:

MICHELLE L. ZBYLUT, Ph.D.
Director

Research accomplished under contract
for the Department of the Army

University of Virginia, Biocomplexity Institute & Initiative, Social and Decision Analytics Division

Technical reviews by:

Dr. Gerald Goodwin, U.S. Army Research Institute

Concurrence for release from ARD received on September 22, 2020.

DISPOSITION

This Technical Report has been submitted to the
Defense Information Technical Center (DTIC).

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
April 2021	Final	January 2020 – October 2020

4. TITLE AND SUBTITLE
An Examination of the Longitudinal Stability of Psychological Measures Contained within the U.S. Army's Global Assessment Tool

5a. CONTRACT NUMBER
W911NF1920164

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
602785A

6. AUTHORS
Ratcliff, Nathaniel J., Thurston, Joel, Goldstein, Joshua R., Lancaster, Vicki A., Shipp, Stephanie S., Keller, Sallie, and Ervin, Kelly S.

5d. PROJECT NUMBER
790

5e. TASK NUMBER
984

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of Virginia, Biocomplexity Institute and Initiative,
Social and Decision Analytics Division
1100 Wilson Blvd, Suite 2910, Arlington, VA 22209

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Institute
for the Behavioral & Social Sciences
6000 6TH Street (Bldg. 1464 / Mail Stop 5610)
Fort Belvoir, VA 22060-5610

10. SPONSOR/MONITOR'S ACRONYM(S)
ARI

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
Technical Report 1397

12. DISTRIBUTION/AVAILABILITY STATEMENT:
Distribution Statement A: Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES
ARI Research POC: Dr. Kelly S. Ervin, Program, Budget, and Strategy Office

14. ABSTRACT

As a component of the U.S. Army's Comprehensive Soldier and Family Fitness program (CSF2), the Global Assessment Tool (GAT) represents a multidimensional constellation of measures designed to assess characteristics related to resilience. Using a foundation of validated measures from prior research, the GAT has been the vehicle for self-assessment to provide Soldiers, their families, and Army Civilians snapshots of their psychosocial wellness. Despite the long history of the measurement instrument (first implemented in 2009) and widespread use (mandatory for all active-duty Soldiers annually), the longitudinal capabilities of the GAT has received little attention.

15. SUBJECT TERMS
Psychological measurement; Longitudinal stability; Resilience; Test battery; Army

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Kelly S. Ervin
Unclassified	Unclassified	Unclassified	Unlimited Unclassified	74	19b. TELEPHONE NUMBER 703-712-3046

Technical Report 1397

An Examination of the Longitudinal Stability of Psychological Measures Contained within the U.S. Army's Global Assessment Tool

Nathaniel J. Ratcliff
Joel Thurston
Joshua R. Goldstein
Vicki A. Lancaster
Stephanie S. Shipp
Sallie Keller
University of Virginia,
Biocomplexity Institute and Initiative,
Social and Decision Analytics Division

Kelly S. Ervin
U.S. Army Research Institute

April 2021

ACKNOWLEDGMENT

The authors thank Kizzann Ratcliff (The Pennsylvania State University), Jason Manley (Research Facilitation Laboratory), Brandon Baze (Research Facilitation Laboratory), Yves Rosseel (Ghent University) and the Army Resilience Directorate for their technical support.

AN EXAMINATION OF THE LONGITUDINAL STABILITY OF PSYCHOLOGICAL MEASURES CONTAINED WITHIN THE U.S. ARMY'S GLOBAL ASSESSMENT TOOL

EXECUTIVE SUMMARY

Research Requirement:

As a component of the U.S. Army's Comprehensive Soldier and Family Fitness program (CSF2), the Global Assessment Tool (GAT) represents a multidimensional constellation of measures designed to assess characteristics related to resilience. Using a foundation of validated measures from prior research, the GAT has been the vehicle for self-assessment to provide Soldiers, their families, and Army Civilians snapshots of their psychosocial wellness. Despite the long history of the measurement instrument (first implemented in 2009) and widespread use (mandatory for all active-duty Soldiers annually), the longitudinal capabilities of the GAT has received little attention. Although prior research has provided an abundance of evidence that GAT measures are psychometrically sound at a given point in time, the longitudinal stability of these measures across time has not yet been examined.

This investigation of the GAT is a by-product of our research on leveraging Department of Defense (DOD) administrative data to model and optimize individual and team performance. While a specific focus on the GAT was unintended, our research on developing predictive models of performance provided a serendipitous opportunity to discover the utility of the GAT as an administrative data source in our research. Given that the GAT was designed to provide a snapshot of a respondent's current levels of resilience, it seemed to be a fitting source to enhance our performance models. We focus on testing whether these measures show variability across multiple time occasions. For our purposes, it is important to establish whether the measurement instrument is sensitive to changes in states that are expected to shift over time—an assumption that has not been empirically tested before.

Procedure:

The GAT data were linked and analyzed within the Person-Event Data Environment (PDE). The PDE is a secure, remote-access, virtual data enclave that allows for housing Army data pertaining to psychological measures, performance indicators, medical information, and administrative personnel records across the careers of individual Soldiers. We had access to GAT data from 2009 to 2017 which covered the time frame for both GAT version 1.0 and 2.0. The sample of respondents for the two studies consisted of 95,277 respondents in Study 1 (GAT 1.0) and 57,771 respondents in Study 2 (GAT 2.0). Across both studies, respondents were active-duty Soldiers who were initially enlisted in the U.S. Army and provided consent to have their data used for study. Respondents were selected with an accession date (i.e., service start date) falling between the start of 2009 and the end of 2014 in Study 1 and between the start of 2012 and the end of 2017 in Study 2. These dates fall within the time period of the GAT 1.0 that was administered Army-wide from 2009 to 2014 and the GAT 2.0 that was administered Army-wide from 2013 to 2017 (GAT 2.0 data were only available through 2017 at the time of our analyses). For each respondent, we selected the first five occasions in which they completed the GAT and

required each respondent have at least two occasions with completed data. Once data had been linked, a triangulation of statisitcal procedures were conducted to assess the longitunidal stability of the GAT including: (a) repeated measues analysis of variance (RM-ANOVA); (b) repeated measures structural equation models (RM-SEM); (c) tests of measurement invariance (MI); and, (d) repeated measures confirmatory factor analysis (RM-CFA).

Findings:

Across statistical approaches, we found the measures on the GAT to be relatively stable over time. However, a few measures such as Organizational Trust, Work Engagement, and Life Meaning did show some meaningful variation across time.

Utlizatiation and Dissemnation of Findings:

This evidence contributes to knowledge of how best to improve the GAT for future use, to which we offer two suggestions. First, the measures contained within the GAT could be modified from their current form to be sensitive to time-varying changes to meet the original intent of an annual self-assessment. The pay-off for the Army is a self-assessment tool that is more effective and efficient. Alternatively, using the current version of the GAT, the frequency in which respondents complete the self-assessment could be shifted from annually to a one-time assessment. This approach could save Soldiers valuable time taking a measure at a single time point upon accessing to the Army.

AN EXAMINATION OF THE LONGITUDINAL STABILITY OF PSYCHOLOGICAL MEASURES CONTAINED WITHIN THE U.S. ARMY'S GLOBAL ASSESSMENT TOOL

CONTENTS

LIST OF TABLES

LIST OF FIGURES

An Examination of the Longitudinal Stability of Psychological Measures Contained within the U.S. Army's Global Assessment Tool

Introduction

The U.S. Army is one of the largest organizations in the world with an active-duty population of 483,941 as of September 2019 (Department of Defense, 2018a). To maintain military readiness and serve at peak performance, the physical, social, and psychological health of Soldiers is of vital importance to mission success. To that end, the Army has endeavored to focus initiatives on promoting and instilling resilience in Soldiers (Casey, 2011; Department of the Army 2014). Underlining this sentiment, the 2018 National Defense Strategy emphasized the need for fielding "a lethal, resilient, and rapidly adapting Joint Force" (Department of Defense, 2018b, p. 1). Thus, fostering psychological resilience in Soldiers, Family members, and Army Civilians is a central aspect of maintaining current and future Army readiness.

The Comprehensive Soldier and Family Fitness Program (CSF2)

In 2008, the Army was facing a critical juncture in time. As a result of prolonged engagements in two major conflicts in Afghanistan and Iraq, Soldiers were undergoing multiple deployments and increased lengths of separations from their families at home. This high operational tempo was putting a difficult strain on the mental, physical, and social health of Soldiers and their families. In response to alarming rates of suicide and post-traumatic stress disorder, Army leadership recognized a need for a program that went beyond the traditional approaches that were largely reactive in focus to adverse events. This led to the idea for a proactive program that provided preventative capabilities through self-assessment, resiliency skill development, and training (Casey, 2011; Cornum et al., 2011). Born out of this need was the Comprehensive Soldier and Family Fitness (CSF2) program. The CSF2 was designed from the start with the aim of providing Soldiers with the skills needed to face and mitigate adversity. Rather than a trait-like construct (cf. Steyer et al., 2015), the CSF2 program conceived resilience as a capacity that could be enhanced through learning and training which allowed for continuous improvement as monitored from self-assessments (Cacioppo et al., 2015; Casey, 2011; Department of the Army, 2010; Vie et al., 2016). The CSF2 program can be broken down into four key components: (a) an online self-assessment of resiliency in the areas of emotional, social, family, and spiritual fitness (see GAT below); (b) a series of tailored self-development modules to help build skills to face adversity, guided by responses on the self-assessment; (c) a course to train 'master resilience trainers' at the unit-level (e.g., battalion) to provide leaders the skills needed to train resiliency in their subordinates; and (d) the introduction of resilience training at every major leader development school in the Army (Casey, 2011; Cornum et al., 2011). Each component was designed to work together to monitor, build, and maintain physical, psychological, and social strengths to provide Soldiers with the ability to serve at their best. We now turn to the component of focal interest, the measurement tool used for self-assessment of resilience capacities.

The Global Assessment Tool (GAT)

Developed by a committee of experts from the Army, academia, and the private sector (Peterson et al., 2011), the Global Assessment Tool (GAT) was designed to serve as the conduit for self-assessment of resilience-related characteristics. Using a theoretical framework grounded in positive psychology (Seligman & Csikszentmihalyi, 2000), the GAT was designed to provide Army personnel and their Families a snapshot of individualized characteristics to create awareness and serve as a basis for improvement in future trainings or learning modules (Peterson et al., 2011; Vie et al., 2016). The GAT was not intended to be used as a selection tool, nor used in promotion or disciplinary actions, or as an indicator of clinical levels of psychological functioning (Cornum et al., 2011; Lester et al., 2015). The GAT was meant to be taken online in about 15 minutes (Lester et al., 2015) and completed no less than once a year for all active-duty Soldiers and deploying Army Civilians (Department of the Army, 2014). For non-deploying Army Civilians and Family members, the GAT was optional but highly encouraged.

Four important psychological fitness components were identified by the committee charged with developing the GAT (Peterson et al., 2011). These components included emotional fitness, family fitness, social fitness, and spiritual fitness (see Table 1). *Emotional fitness* (e.g., adaptability, character, depression) reflects one's mood, satisfaction, coping styles, and character strengths. *Social fitness* (e.g., loneliness, organizational trust) is indicative of one's feelings towards close friends, unit members, and leaders. *Family fitness* (e.g., family satisfaction and support) is an indicator of one's personal relationships with family and romantic partners. *Spiritual fitness* (i.e., life meaning) reflects the degree to which one ascribes meaning and purpose to their life and world around them. It is important to note that the term 'spiritual' is not indicative of any sort of religiosity or religious belief but an existential meaning one draws from their life and the world around them (Peterson et al., 2011). A fifth component, physical fitness, was added later in version 2.0 in 2014 which assess aspects of nutrition, sleep habits, and substance abuse (Lester et al., 2015).

Drawn from the existing literature, the majority of items (~90%) used in the GAT were taken from well-validated measures (Peterson et al., 2011). For instance, a portion of the expanded form of the Positive and Negative Affect Schedule (PANAS-X; Watson & Clark, 1999; Watson et al., 1988) was used as a measure of affect for the emotional fitness component. For constructs that did not have pre-existing validated scales, measures were developed by the research team (Peterson et al., 2011; Vie et al., 2016) to include in the assessment tool (e.g., adaptability, family support, friendship). Each measurement scale on the GAT represented a sub-scale of one of the four fitness components. Measures were worded to provide a state-like measurement of constructs. For instance, the stem "In the past four weeks…" was used for many of the items as the anchor point for judgments. Taken together, the GAT measures were intended to provide a gauge psychosocial fitness as a self-awareness tool for Soldiers taking the assessment. For a detailed description of GAT measures, please see *Appendix*. Of note, as of 2019, the Army currently refers to the GAT as the Azimuth Check within the ArmyFit program.

Table 1

Summary of Measurement Components within the GAT by Version

Component	Component Measures	GAT Version: GAT 1.0 (2009–2014)	GAT Version: GAT 2.0 (2013–2017 †)	Brief Description	Sample Item	Source
Emotional Fitness		77 items	64 items			
	Adaptability	3 items	3 items	Ability to alter one's course and perceived cognitive flexibility.	"I am good at changing myself to adjust to changes in my life." [a]	Inspired by prior research (Martin & Rubin, 1995)
	Active Coping (Problem-Focused)	5 items	5 items	Active coping strategies that involve planning or taking directed action.	"When something stresses me out, I try to solve the problem." [a]	Adapted from Brief COPE (Carver, 1997; Carver et al., 1989)
	Passive Coping (Emotion-Focused)	3 items	3 items	Passive coping strategies that involve venting or displacement and disengagement.	"I usually keep my emotions to myself." (reverse-coded) [a]	Adapted from Brief COPE (Carver, 1997; Carver et al., 1989)
	Character	24 items	18 items	Character strengths within the virtues of wisdom, courage, humanity, justice, temperance, and transcendence.	"Critical thinking." [b]	Adapted from Character Strengths Test (Peterson, 2007; Peterson & Seligman, 2004)
	Catastrophizing	7 items	3 items	Internal explanatory style of attributions towards negative events.	"When bad things happen to me, I expect more bad things to happen." [a]	Adapted from the Attributional Style Questionnaire (Peterson et al., 1982; Peterson et al., 2001)
	Depression	10 items	10 items	Prevalence of depressive symptoms of feeling down, depressed, or hopeless.	"Feeling down, depressed, or hopeless." [c]	Adapted from the Patient Health Questionnaire (Kroenke et al., 2001; Spitzer et al., 1999)
	Optimism	4 items	4 items	Generalized expectation for positive future events.	"In uncertain times, I usually expect the best." [d]	From the Life Orientation Test (Scheier & Carver, 1985; Scheier et al., 1994)
	Positive Affect	10 items	9 items	Subjective feelings of positive affect.	"Inspired." [e]	From PANAS-X (Watson & Clark, 1999; Watson et al., 1988)
	Negative Affect	11 items	9 items	Subjective feelings of negative affect.	"Upset." [e]	From PANAS-X (Watson & Clark, 1999; Watson et al., 1988)
Family Fitness		5 items	5 items			
	Family Satisfaction ‡	2 items	2 items	Satisfaction with family and romantic relationships.	"How satisfied are you with your family?" [f]	Directorate of Basic Combat Training's Experimentation and Analysis Element, Fort Jackson
	Family Support ‡	3 items	3 items	Perception that family supports one's service and that the Army supports one's family.	"My family supports my decision to serve in the Army." [d]	Directorate of Basic Combat Training's Experimentation and Analysis Element, Fort Jackson
Social Fitness		18 items	15 items			
	Friendship ‡	6 items	4 items	Degree to which there are people for whom one can depend on for support when needed.	"I have someone to talk to when I feel down." [g]	Original items
	Loneliness	3 items	3 items	Feelings of being alone and separated from others.	"How often do you feel left out?" [e]	Adapted from the UCLA Loneliness Scale (Russell et al., 1978)
	Organizational Trust	5 items	4 items	Trust in the organization (i.e., peers, leaders) in terms of ability, benevolence, and integrity.	"I trust my fellow Soldiers in my unit to look out for my welfare and safety." [d]	Inspired by prior research (Mayer & Davis, 1999; Mayer et al., 1995; Sweeney et al., 2009)
	Work Engagement	4 items	4 items	Satisfaction and commitment to work.	"My work is one of the most important things in my life." [a]	Adapted from the Work as a Calling Scale (Wrzesniewski et al., 1997) and the Orientations to Happiness Scale (Peterson et al., 2005)
Spiritual Fitness		5 items	5 items			
	Life Meaning	5 items	5 items	Sense of purpose and meaning to life and work.	"I believe there is a purpose for my life." [a]	Adapted from the Brief Multidimensional Measure of Religiousness/Spirituality (Fetzer Institute, 1999) and Purpose in Life Scale (Crumbaugh, 1968)
Physical Fitness		NM	74 items			
	Physical Activity ‡	NM	12 items	Degree and types of physical activity.	"How many days per week did you perform the vigorous activity in the last 30 days?" [g]	Various DOD questionnaires
	Nutrition ‡	NM	19 items	Food intake and use of nutritional supplements.	"Do you take dietary supplements?" [g]	Various DOD questionnaires
	Lifestyle Behaviors ‡	NM	25 items	Sleep habits, alcohol consumption and tobacco use.	"How would you rate your satisfaction with your sleep?" [g]	Various DOD questionnaires
	Risk Factors ‡	NM	18 items	Family history of health problems and willingness to engage in risky behavior.	"How often do you text while driving?" [g]	Various DOD questionnaires

Note. GAT = Global Assessment Tool; † GAT remains in operational use (under new name Azimuth Check), but data available for the current research ended in 2017; ‡ Measure not examined in the current research; NM = not measured in GAT version. [a] 5-point scale (1 = *Not like me at all*, 5 = *Very much like me*). [b] 11-point scale (0 = *Never*, 10 = *Always*). [c] 5-point scale (1 = *Not at all*, 5 = *Every day*). [d] 5-point scale (1 = *Strongly disagree*, 5 = *Strongly agree*). [e] 5-point scale (1 = *Never*, 5 = *Most of the time*). [f] 5-point scale (1 = *Not at all satisfied*, 5 = *Extremely satisfied*). [g] Various response options.

Prior Research Findings with the GAT

Though early on the GAT faced some criticism for a lack of empirical testing (Eidelson et al., 2011; Steenkamp et al., 2013), the current state of research on the GAT shows the instrument to be psychometrically sound (Lester et al., 2015; Vie et al., 2016). The GAT has been shown to be internally-consistent with reliability estimates that exceed the benchmark of .70 for most measurement scales (Lester et al., 2015; Peterson et al., 2011; Shingleton et al., 2016; Vie et al., 2016) and is structurally-sound as indicated by well-defined and well-fitting factor structures (Shingleton et al., 2016; Vie et al., 2016). In addition, the GAT has also been found to be relatively invariant between certain sub-populations in the Army (e.g., age and rank) and the GAT measures show divergent and convergent validity with one another (Vie et al., 2016). In sum, the preponderance of evidence to date suggests the GAT to be a reliable and valid measurement instrument in the Army population.

Above and beyond its utility as a psychometrically-sound instrument, the GAT has also been found to have predictive value with many outcomes of interest to the CSF2 program. Researchers have found that the GAT is predictive of first-term attrition; Soldiers with poor well-being scores (e.g., depression, positive affect, adaptability) were found to be more likely to attrit from military service (Cunha et al., 2015). Similarly, poor GAT scores on the emotional and social dimension have been linked to higher propensities of dishonorable discharge from the service (Shingleton et al., 2016) and increased rates of suicide and violent offenses (Lester et al., 2011). By contrast, more positive scores on GAT measures (e.g., optimism, catastrophizing, work engagement) have been associated with individuals in elite occupational roles like the U.S. Rangers (Lester et al., 2015) and the life meaning measure has been associated with a greater likelihood of reenlistment for another term in the Army (Shingleton et al., 2016).

Current Research

Our research aims to examine the longitudinal stability of measures contained within the GAT. This investigation of the GAT is a by-product of our research on leveraging Department of Defense (DOD) administrative data to model and optimize individual and team performance. The Army possesses a trove of administrative data (e.g., personnel records, training scores) and it has been recommended these data be explored for use to support mission effectiveness (National Academies of Sciences, Engineering, and Medicine, 2017, 2019). While a specific focus on the GAT was unintended, our research on developing predictive models of performance provided a serendipitous opportunity to discover the utility of the GAT as an administrative data source in our research. Given that the GAT was designed to provide a snapshot of a respondent's current levels of resilience, it seemed to be a fitting source to enhance our performance models. Building upon prior work that has investigated the psychometric properties of the GAT measures at a single time occasion (e.g., Cunha et al., 2015; Lester et al., 2011; Vie et al., 2016), we focus on testing whether these measures show variability across multiple time occasions. For our purposes, it is important to establish that the measurement instrument is sensitive to changes in states that are expected to shift over time—an assumption that has not been empirically tested before.

The current research offers several key strengths to test our research question. First, we examine the longitudinal stability of GAT measures across five time occasions using both versions of the GAT (GAT 1.0 in Study 1 and GAT 2.0 in Study 2). Second, we can use very large sample sizes that afford maximal statistical power to detect even small effects. Third, we use a multi-method approach based on several different statistical procedures to triangulate our findings at the composite scale level (e.g., analysis of variance and structural equation models) and the latent factor level (e.g., measurement invariance and confirmatory factor analysis).

The findings of this research provide evidence that can inform Army practitioners on the longitudinal properties of the GAT. We demonstrate the GAT to be invariant over time, suggesting that costs could be reduced by fielding the tool at a single time point (e.g., upon entry into the Army) and/or the GAT measures could be re-aligned with future development to ensure that measures capture time-varying aspects of respondents. This research also serves academic researchers and research organizations who wish to use the GAT in their modelling of DOD data; understanding the psychometric properties regarding scale reliability, measurement invariance, and longitudinal change can help inform whether to treat GAT measures as time-varying variables or trait-like predictor variables in models.

Method

Sample Selection Criteria

For our analyses, we had access to GAT data from 2009 to 2017. The sample of respondents for the two studies consisted of 95,277 respondents in Study 1 (GAT 1.0) and 57,771 respondents in Study 2 (GAT 2.0). Across both studies, respondents were active-duty Soldiers who were initially enlisted in the U.S. Army and provided consent to have their data used for study. Although we were unable to directly assess the GAT data of non-consenting respondents, we were able to analyze certain characteristics of these respondents and found that they were very similar to consenting respondents in terms of age, gender, race, and job specialty (see Table 2). Using a cross-sequential longitudinal design with planned missing data (Little, 2013, p. 38), respondents were selected with an accession date (i.e., service start date) falling between the start of 2009 and the end of 2014 in Study 1 and between the start of 2012 and the end of 2017 in Study 2. These dates fall within the time period of the GAT 1.0 that was administered Army-wide from 2009 to 2014 and the GAT 2.0 that was administered Army-wide from 2013 to 2017. To our knowledge, the GAT 2.0 is still in operational use but data after 2017 were not available at the time of our analyses. For each respondent, we selected the first five occasions in which they completed the GAT and required each respondent have at least two occasions with completed data. To ensure the GAT 2.0 sample in Study 2 was independent from the GAT 1.0 sample in Study 1, respondents were removed from the sample in Study 2 if data were present for them in Study 1.

Model Testing Strategy

All data were accessed and analyzed in the Army's Person-Event Data Environment (PDE) using the statistical software R 3.6.1 (see *Appendix* for example code). The PDE is a secure, remote-access, virtual data enclave that allows for housing Army data pertaining to

psychological measures, performance indicators, medical information, and administrative personnel records across the careers of individual Soldiers (National Academies of Sciences, Engineering, and Medicine, 2017; Vie et al., 2015). The PDE is supported by the Army and is administered by the Research Facilitation Laboratory (RFL) which is overseen by the Army Analytics Group (AAG). Use of the PDE is available to researchers and institutions that wish to conduct research using Army and DOD data sources provided proper approvals are met (Knapp et al., 2018). Prior to conducting analyses, approvals for the study protocol were gained from the University of Virginia Institutional Review Board and the Army Human Research Protections Office.

To provide a comprehensive examination of the research question as to whether GAT measures were stable over time, we analyzed the data using four statistical approaches: repeated measures analysis of variance (RM-ANOVA), repeated measures structural equation model (RM-SEM), tests of measurement invariance (MI), and repeated measures confirmatory factor analysis (RM-CFA). The first two of these approaches analyze the data at the composite scale level (RM-ANOVA, RM-SEM); that is, examining the measures when the individual items are averaged together into a scale. The last two approaches analyze the data at the latent construct level using the individual item indicators and the corresponding latent factors the items reflect (MI, RM-CFA). Each of these approaches (discussed in detail in *Appendix B*, for conceptual depictions of models, see *Appendix C*) has its own strengths and weakness. For an examination of individual-level (vs. mean-level) stability, see *Appendix I, J,* and *K*. Considered together, our approaches help triangulate evidence to create a comprehensive picture of the longitudinal stability of GAT measures. Given that there were multiple tests across the 13 measures, a study-wide statistical correction factor was taken into account for the significance tests.

Given the large sample sizes, we used measures of effect size to examine the magnitude of the effect (cf. Steenkamp et al., 2013) of the omnibus tests (i.e., testing the null assumption that means for scale and latent variables do not differ across time occasions). Across analytic approaches, generalized eta-squared (η^2) was used as a measure of effect size across approaches for comparison purposes along with 90% Confidence Intervals (Bakeman, 2005; Lakens, 2013; Steiger, 2004). Cohen's (1988) suggested guidelines for η^2 interpretation were as follows: .02 = small, .13 = medium, .26 = large; these benchmarks are somewhat arbitrary but in the absence of prior research or indications of meaningful change, these benchmarks were used. In addition, for model comparisons using a chi-square difference (i.e., likelihood ratio) test (i.e., RM-SEM, MI, RM-CFA), Cohen's ω was used which has the following magnitude interpretation: 0.10 = small, 0.30 = medium, 0.50 = large. For details regarding the calculations of these effect size indices, see *Appendix B*.

Results

Sample and Data Characteristics

A breakdown of the sample characteristics for both studies can be found in Table 2. Overall, the respondent sample size in Study 1 was 95,277 Soldiers with 4,968 having completed data for all five time occasions and, in Study 2 was 57,771 Soldiers with 679 having completed data for all five time occasions. Sample characteristics did not differ depending on the number of

time occasions completed by respondents. Respondent age, gender, and self-identified race were all consistent with DOD reports on demographics for respondents serving in the U.S. Army (Department of Defense, 2015). Consistent with administration timelines for the GAT, the average duration between time occasions was close to the year mark at 414 days (*SD* = 152.01 days) in Study 1 and 398 days (*SD* = 154.04 days) in Study 2.

Table 2

Sample Characteristics for Study 1 (GAT 1.0) and Study 2 (GAT 2.0) with Comparison to Non-Consenters

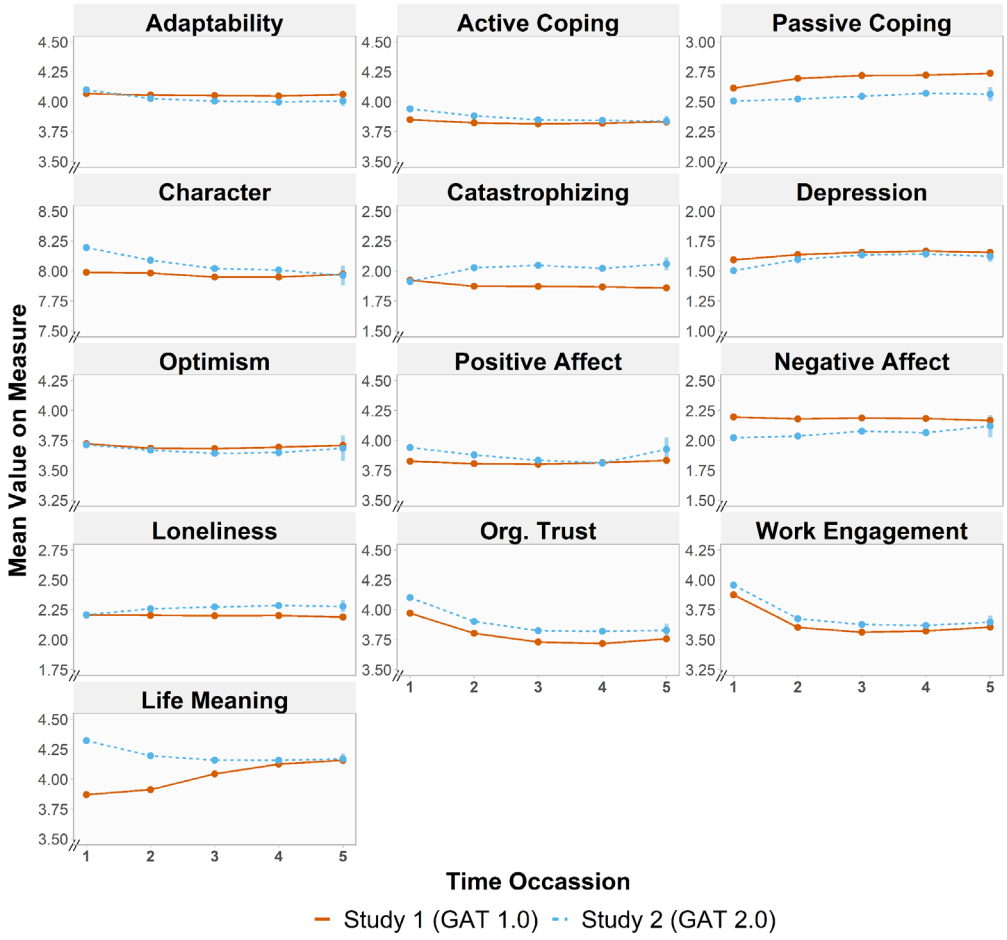
		GAT Version		
Sample Characteristics		GAT 1.0 (Study 1)	GAT 2.0 (Study 2)	Non-Consenters
Maximum Sample Size (*N*)				
	Study Sample Size	95,277	57,771	179,009
Time Occasion Completions (*n*)				
	T1	95,277	57,771	—
	T2	95,277	57,771	—
	T3	53,894	19,194	—
	T4	23,274	4,722	—
	T5	4,968	679	—
Time Occasion Characteristics (Mean/SD)				
	M_{Diff} (*SD*)	414.18 (152.01)	398.79 (154.04)	—
	M_{Age} (*SD*)	23.34 (4.80)	22.46 (3.62)	22.03 (3.73)
Military Occupational Specialty Type (%)				
	Combat Arms	32.15	28.41	38.99
	Combat Support	29.96	29.33	25.88
	Combat Service Support	38.43	38.18	35.13
Gender (%)				
	Male	84.43	84.03	83.44
	Female	15.57	15.97	16.56
Race & Ethnicity (%)				
	Caucasian	59.48	51.90	59.47
	African-American	18.44	23.38	18.90
	Hispanic	12.95	16.71	13.90
	Asian	3.10	6.60	3.93
	Native Hawaiian	0.92	0.17	0.55
	Native Indian	0.68	0.67	0.78
	Mixed Race/Other	4.39	0.56	2.47

Note. GAT = Global Assessment Tool; M_{Diff} = mean number of days between time occasions; M_{Age} = mean age of respondent at first time occasion or at accession to the Army for non-consenters. The non-consenter group represents those who specifically did not consent to have their GAT data used for research purposes. However, research-approved administrative data for these individuals (outside the GAT data) were available to examine their similarities to the samples that did consent their data for GAT research. In addition, overall sample demographic frequencies reported here were similar for each progressive group of respondents who had complete data for three, four, and five time occasions, usually within less than one percentage point. This finding suggests that subsamples were relatively well-representative of the larger samples for each study.

A summary of the descriptive statistics and reliabilities for the composite scale measures can be found in the *Appendix* (Table D1 and Table D2). A plot of the means for each measure across time occasions can be found in Figure 1. Across GAT measures for both studies, bivariate correlations between time occasions were significant (all *p*s < .001). In addition, measures of internal consistency using omega total (for a review on the usage of omega total over Cronbach's alpha, see McNeish, 2018) were all above the .70 threshold across GAT measures and time occasions, the exception of adaptability in Study 2 which had omega total values ranging from 0.69–0.74.

Figure 1

Mean Plot of Global Assessment Tool (GAT) Measures Across Five Time Occasions



Note. Mean plot of Global Assessment Tool (GAT) measures across five time occasions for Study 1 representing GAT version 1.0 (solid orange line) and Study 2 representing GAT version 2.0 (dotted blue line). Data represent respondents with no missing time occasions. Error bars represent 95% confidence intervals. Data are plotted on a truncated y-axis depicting a one-point range for each scale.

Repeated Measures Analysis of Variance (RM-ANOVA)

The detailed results of the RM-ANOVA analyses are presented in the *Appendix*, Tables E1 and E2 and a summary of effect size comparisons are presented in Figure 2. Nearly all the measures had significant *F*-tests except for the adaptability measure in Study 1 and the measures of loneliness, optimism, positive affect, and negative affect in Study 2. However, when examining the effect size of these tests, only the life meaning measure exceeded the threshold of a small effect in Study 1 and organizational trust in Study 2. All other effects were borderline small (i.e., organizational trust, work engagement) or were effectively zero. When looking at the pattern of means (see Figure 1), organizational trust and work engagement showed a downward trend after the first measured occasion across both studies. By contrast, life meaning showed an upward trend over the five time occasions in Study 1 but a downward trend in Study 2.

Repeated Measures Structural Equation Model (RM-SEM)

The results of the RM-SEM analyses are presented in *Appendix*, Tables F1 and F2 and effect sizes are presented in Figure 2. All the chi-square difference tests between the full and null models for the measures were significant (ps < .05 in Study 1 and ps < .001 in Study 2) which is to be expected given that the chi-square is overly sensitive to small, unimportant deviations from perfect model fit when sample sizes are large (Chen, 2007; Putnick & Bornstein, 2016). In Study 1, the magnitude of change in the comparative fit index (ΔCFI) for the measures of passive coping, catastrophizing, depression, life meaning, organizational trust, and work engagement all exceeded the recommended −.01 benchmark (Chen, 2007; Cheung & Rensvold, 2002; Marsh et al., 2010; Putnick & Bornstein, 2016) while life meaning, organizational trust, and work engagement exceeded the recommended magnitude of change in root mean square error of approximation (ΔRMSEA) of .015 (Chen, 2007). In Study 2, the magnitude of change in CFI for nearly all measures (except passive coping and negative affect) exceeded the −.01 benchmark while only organizational trust and work engagement had changes in RMSEA which exceeded .015. Similarly, when examining the measures of effect size for the chi-square difference tests in both studies, only life meaning, organizational trust, and work engagement exceeded the threshold of a small effect for both ω and η^2, all other measures were trivial or near zero. For analogous results allowing for random slopes and intercepts, see *Appendix*.

Measurement Invariance (MI)

The MI analysis results are presented in the *Appendix*, Tables G1 and G2. In the sub-sections that follow, results are presented for each invariance type: configural, weak, and strong.

Configural Invariance

Configural invariance was accessed by evaluating how well the specified model (i.e., pattern of free and fixed loadings across time occasions) fit the observed data. In general, configural models for the measures were relatively well-fitting. Moreover, although all the chi-square tests were significant across studies (ps < .001), the alternative fit indices were largely within the recommended ranges (CFIs > .90, TLIs > .90, RMSEAs < .08, standardized root mean square residuals < .09; see Hu & Bentler, 1999) with some measures like character (Study 1: CFI = .846; Study 2: CFI = .865), negative affect (Study 1: CFI = .850; Study 2: CFI = .896), and optimism (Study 2: CFI = .832) falling slightly outside the range on the CFI but not on the other indices (e.g., RMSEAs < .05).

Weak (Metric) Invariance

To test weak invariance, a model was estimated for each measure where factor loadings were set (i.e., constrained or restricted) to be equivalent across all five time occasions. This model was then compared to the previously estimated configural model for the respective measure using a chi-square difference test. In general, comparisons between the two nested models provided supporting evidence for weak invariance across measures. Moreover, although all the chi-square tests were significant across studies (ps < .001), the relative change in

magnitude of the alternative fit indices were all within the recommended ranges (e.g., ΔCFI < −.01, ΔRMSEA < .015) or showed improvements in fit. Similarly, when examining the measures of effect size for the chi-square difference tests, none of the measures exceeded the threshold of a small effect for both ω and η^2, all other measures were trivial or near zero.

Strong (Scalar) Invariance

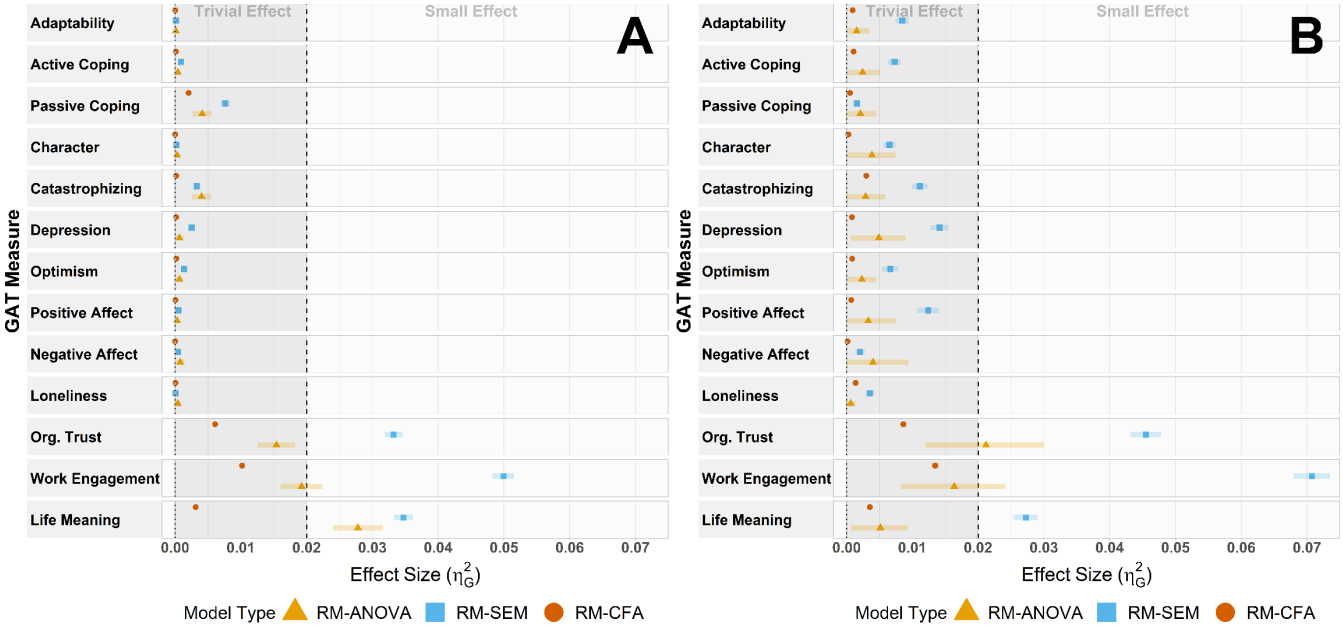
To test strong invariance, a model was estimated for each measure where item intercepts were set to be equivalent across all five time occasions. This model was then compared to the previously estimated weak invariance model for the respective measure using a chi-square difference test. In general, comparisons between the two nested models provided supporting evidence for strong invariance across measures. Moreover, although all the chi-square tests were significant across studies (*p*s < .001), the alternative fit indices were largely within the recommended ranges with the exception of the measures life meaning (Study1: ΔCFI = −.062, ΔRMSEA = .014; Study 2: ΔCFI = −.013, ΔRMSEA = .003), organizational trust (Study 1: ΔCFI = −.016, ΔRMSEA = .003), and negative affect (Study 2: ΔCFI = −.011, ΔRMSEA = .002) which had mixed changes in fit. Similarly, when examining the measures of effect size for the chi-square difference tests, only life meaning exceeded the threshold of a small effect in Study 1 for both ω and η^2, all other measures were trivial or near zero.

Repeated Measures Confirmatory Factor Analysis (RM-CFA)

The results of the RM-CFA analyses are presented in the *Appendix*, Tables H1 and H2 and effect sizes are presented in Figure 2. All the chi-square difference tests between the full and null models for the measures were significant (*p*s < .01 in Study 1 and *p*s < .001 in Study 2). In addition, the magnitude of change in CFI (ΔCFI) for the measures of organizational trust and work engagement exceeded the −.01 benchmark. None of the measures exceeded the .015 benchmark for magnitude of change in RMSEA (ΔRMSEA). When examining the measures of effect size for the chi-square difference tests, only life meaning, organizational trust, and work engagement met or exceeded the threshold of a small effect for ω, but none crossed the small threshold for η^2. Effect size indicators for all other measures were trivial or near zero.

Figure 2

Plot of Effect Sizes for Global Assessment Tool (GAT) Measures Across Five Time Occasions



Note. RM-ANOVA = repeated measures analysis of variance, RM-SEM = repeated measures structural equation model, RM-CFA = repeated measures confirmatory factor analysis. Plot of effect sizes for Global Assessment Tool (GAT) measures across five time occasions for Study 1 representing GAT version 1.0 (Panel A) and Study 2 representing GAT version 2.0 (Panel B). The shaded area right of the dotted line at 0.00 represents no effect and the values to the right of the dashed line at 0.02 represent values considered to be in the small effect range according to Cohen (1988).

General Discussion

The current research presented a longitudinal examination of the GAT. Across two studies, our findings support that the majority of psychological measures assessed within the GAT are stable over time. Moreover, the use of large samples of active-duty Army Soldiers in conjunction with a triangulated analytic approach provides robust evidence for these findings.

Measure Reliability and Invariance Over Time

When considering internal reliability, our findings suggest that measures within the GAT are reliable. Estimates of internal reliability of the 13 measures met or exceeded the recommended threshold at the first time occasion and all ensuing time occasions, thus extending the evidence of prior work that examined that GAT at a single time occasion (cf. Vie et al., 2016). In addition, we also find evidence that the GAT measures possess relatively good measurement invariance across time. As indicated by measures of fit and effect size, GAT measures seem to maintain a similar factor structure (i.e., configural invariance), have similar item loadings (weak/metric invariance), and similar item intercepts (strong/scalar invariance) across time occasions. This evidence helps rule out the possibility that mean differences at the

scale or latent factor level are due to other aspects of measurement unrelated to mean differences due to the transition of time.

Measure Longitudinal Stability: Scale- and Factor-Level Mean Changes Over Time

The primary objective of the current research was to examine the longitudinal stability of GAT measures. Several analytic techniques were used to test the null hypothesis that the means of measures were equivalent across the five time occasions. These techniques found consistent evidence whether the data were analyzed at the scale-level (i.e., RM-ANOVA, RM-SEM) or factor-level (i.e., RM-CFA). First, most omnibus tests of change over time were significant across measures. This result was not surprising since tests of significance can become sensitive to even minute differences with large sample sizes. Second, when considering the magnitude of effects, all but three measures failed to pass the threshold of a small effect size. Effect sizes for most measures were trivial or near zero in magnitude. For these measures that did not change to a meaningful magnitude, it suggests that they tend to measure more trait-like characteristics of respondents rather than measuring attributes that are sensitive to time-varying states.

Three measures did consistently show some meaningful degree of change over time, albeit small, as indicated by effect size measures. For example, organizational trust and work engagement both showed small effects with similar downward trends after the first measurement occasion. This result suggests that Soldiers felt less trust and commitment to their units and job as time progressed. The measure of life meaning also indicated some change over time but with different patterns from Study 1 to Study 2. In Study 1, an upward trend was observed for all the subsequent time occasions after the first. This result suggested that Soldiers found greater life meaning and purpose as their Army career progressed. By contrast, in Study 2, a downward trend (similar to organizational trust and work engagement) was observed for all the subsequent time occasions after the first which suggested a decrease in life meaning and purpose in the Army as time went on. We do not have a clear explanation for why this reversal was observed, but it may be attributable to a cohort effect. Future research is needed to unravel this trend reversal.

Taken together, our results suggest that the GAT serves as stable, multidimensional measure of Soldier characteristics. Using a triangulated modeling approach, the evidence is consistent at the composite scale level and at the latent factor level. Except for a few measures, all analytic approaches provide evidence that GAT measures are stable (unchanging) over time.

Limitations and Future Directions

Although this research presents a novel examination of GAT measures across time, these findings have a few limitations and raise several interesting questions that warrant future research. First, the current research only examined a sample of active-duty Soldiers who had enlisted in the U.S. Army. We did not examine commissioned officers or those in other components of the Army (e.g., Army Reserve, Army National Guard, Army Civilians, families of active-duty Soldiers). Although past research did not find differences between ranks (i.e., enlisted vs. officer) on the GAT measures (Vie et al., 2016), it would be interesting to examine whether measurement stability fluctuates for officers differently than enlisted members. In the Army, officers take on a unique set of responsibilities in their roles as commanders and leaders

that might shift responses over time on GAT measures differently than their enlisted counterparts (e.g., organizational trust, work engagement). Furthermore, it would be interesting to examine the structure and stability of the GAT measures in a civilian population such as civilians who work for the Army or Family members of active-duty Soldiers. Unlike active-duty Soldiers, civilians are relatively not under the same high operational demands to function at a maximal level that is infused into Army culture. Therefore, civilian respondents may show more variability in their responses to GAT measures of psychological resilience as there may be less cultural pressure to maintain maximum performance.

We did not examine for differences by other demographic characteristics of Soldiers such as gender, race, or age. In Vie and colleagues (2016) cross-sectional analysis, they found evidence that the GAT measures were invariant on age but partially invariant on gender at a single time point. Although we have no specific hypotheses regarding the influence of race and gender on the longitudinal stability of the GAT measures, it is an area that should be investigated to better understand the potential for moderating influences on response patterns.

Another area of future research could examine how the experiences of Soldiers affect the GAT measures. To date, there has been little research looking at how a Soldiers' job type (i.e., Military Occupational Specialty; MOS) or other career events (e.g., deployment, marriage, disciplinary actions) might affect GAT measures. For example, the Army is composed of numerous job types that each serve a unique function in the organizational structure. Some Soldiers serve in combat roles (e.g., infantryman) while others serve in supportive roles that are not often directly in combat or in the field at all (e.g., cyber network defender). Given these unique roles and job functions, research should examine whether the type of jobs Soldiers perform might have any impact on the longitudinal measurement of the GAT. In addition, given the longitudinal analysis, comparing other changes in a Soldiers' career might be fruitful area of future research. For instance, patterns of deployment (i.e., number and frequency) between measurement points of the GAT may prove to be influential.

Practical Implications

The findings from the current research offer several practical implications that are useful to both academic researchers and Army practitioners interested in the administration and/or analytic treatment of the GAT measures. Given that GAT contains measures used by many researchers (e.g., positive/negative affect, loneliness, work engagement, organizational trust), the current research provides robust estimates of these measures in terms of measure structure and longitudinal stability. For example, this work replicates many of the initial test-retest reliability findings on the PANAS-X that were conducted using a few time points separated by a few months or years (Watson & Clark, 1999; Watson & Walker, 1996). For analysts, interested in doing research with GAT data, our findings suggest that many of the current GAT measures could be treated as trait measures at a single time point for predicting outcomes of interest. This should help simplify the modeling of GAT measures by removing the need to model GAT measures as time-varying constructs.

Designing measures that are well-suited to measure change has often been a difficult area for longitudinal research and measurement (Little, 2013). The result that many of the measures

assessed by the GAT are relatively stable over time, suggests that these stable measures could be considered to be measuring more trait-like characteristics of Soldiers than state-like characteristics that vary over time (Steyer et al., 2015; Zukerman, 1983). During the development of the GAT for the CSF2 program, it had been assumed that these measures would serve as status indicators of a Soldiers' psychosocial fitness to provide awareness during their career (see Lester et al., 2011; Peterson et al., 2011). However, from our evidence, it does not seem like many of these measures are meaningfully sensitive to time-varying changes. This finding has important implications for the administration and use of the GAT, to which we provide two mutually exclusive recommendations.

First, the GAT measures could be modified to help re-align the measures to assess state-like changes in Soldier characteristics. For future instantiations of the GAT, we would suggest that the items on the measures be tuned to be more time sensitive. This process might involve any combination of (a) changing the wording of the items to be more specific than general; (b) changing the timeframe of the item stems to be more narrow than wide (e.g., "In the past few days…" vs. "In the past four weeks…"); (c) dropping individual items that are the least time invariant; and/or (d) developing and adding new items that are demonstrated to be sensitive to changing states over time. Modification of GAT measures would help meet the original intent of the GAT to serve as an effective snapshot of psychosocial health and fitness of Soldiers.

Alternatively, if modification is not feasible or desirable, then we suggest that the GAT, in its currently designed form, could be administered at a single time point (e.g., upon joining the Army). A single administration would be sufficient to assess the desired trait-like characteristics given that most measures are time invariant. Administration could occur more than once for measures that do show some variability across time (e.g., life meaning, organizational trust, work engagement) or for measures that might be useful for follow-up after especially undesirable scores (e.g., depression, loneliness) that are linked to important outcomes of interest like suicide and violent crime (see Lester et al., 2011). Furthermore, if the GAT was administered once at the onset of a person's career, this would shift the use of the GAT from a continuous awareness tool to an instrument well-suited for identifying resilience traits in Soldiers joining the Army and could be used to flag those who might benefit from further training. Also, reducing the GAT to a one-time assessment would save time and resources that are presently used to field an annual assessment.

Conclusion

This research adds to the growing literature on the constellation of measures contained within the GAT. Specifically, the current findings provide additional evidence that GAT measures are internally consistent, fit an intended structure, and, for the majority of measures, are longitudinally stable over multiple time occasions. Considering this knowledge, future use of the GAT could focus on modifying the instrument to be more sensitive to state-like changes or using the current measures at a single time assessment. Changes informed by this research could help improve the battery of measures to fulfil the goals set forth by the CSF2 program of providing an instrument that effectively assesses indicators of psychosocial health to maintain resiliency of Army Soldiers, their families, and Army Civilians.

References

Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. *Behavior Research Methods*, *37*, 379–384. https://doi.org/10.3758/BF03192707

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107,* 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Blampied, N. M. (2017). Analyzing therapeutic change using modified Brinley plots: History, construction, and interpretation. *Behavior Therapy*, *48*, 115–127. https://doi.org/10.1016/j.beth.2016.09.002

Breitsohl, H. (2019). Beyond ANOVA: An introduction to structural equation models for experimental designs. *Organizational Research Methods*, *22*, 649–677. https://doi.org/10.1177/1094428118754988

Cacioppo, J. T., Adler, A. B., Lester, P. B., McGurk, D., Thomas, J. L., Chen, H., & Cacioppo, S. (2015). Building social resilience in soldiers: A double dissociative randomized controlled study. *Journal of Personality and Social Psychology*, *109*, 90–105. https://doi.org/10.1037/pspi0000022

Carver, C. S. (1997). You want to measure coping but your protocol's too long: Consider the brief COPE. *International Journal of Behavioral Medicine*, *4*, 92–100. https://doi.org/10.1207/s15327558ijbm0401_6

Carver, C. S., Scheier, M. F., & Weintraub, J. K. (1989). Assessing coping strategies: A theoretically based approach. *Journal of Personality and Social Psychology*, *56*, 267–283. https://doi.org/10.1037//0022-3514.56.2.267

Casey, G. W. (2011). Comprehensive soldier fitness: A vision for psychological resilience in the U.S. Army. *American Psychologist*, *66*, 1–3. https://doi.org/10.1037/a0021930

Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement. *Structural Equation Modeling, 14,* 464–504. https://doi.org/10.1080/10705510701301834

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling, 9,* 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Christensen, L., & Mendoza, J. L. (1986). A method assessing change in a single subject: An alteration of the RC index. *Behavior Therapy*, *17*, 305–308. https://doi.org/10.1016/S0005-7894(86)80060-0

Cohen, J. 1988. *Statistical power analysis for the behavioral sciences*, 2nd Edition. Routledge.

Cornum, R., Matthews, M. D., & Seligman, M. E. P. (2011). Comprehensive soldier fitness: Building resilience in a challenging institutional context. *American Psychologist*, *66*, 4–9. https://doi.org/10.1037/a0021420

Crumbaugh, J. C. (1968). Cross-validation of Purpose-in-Life test based on Frankl's concepts. *Journal of Individual Psychology, 24,* 74–81.

Cunha, J. M., Arkes, J., Lester, P. B., & Shen, Y.-C. (2015). Employee retention and psychological health: Evidence from military recruits. *Applied Economics Letters, 22,* 1505–1510. https://doi.org/10.1080/13504851.2015.1042136

De Fruyt, F., Bartels, M., Van Leeuwen, K. G., De Clercq, B., Decuyper, M., & Mervielde, I. (2006). Five types of personality continuity in childhood and adolescence. *Journal of Personality and Social Psychology*, *91*, 538–552. https://doi.org/10.1037/0022-3514.91.3.538

Department of Defense. (2015). *2015 Demographics: Profile of the military community.* Washington, DC, Headquarters, Department of Defense. Available: https://download.militaryonesource.mil/12038/MOS/Reports/2015-Demographics-Report.pdf [March 2020]

Department of Defense. (2018a). *2018 Demographics: Profile of the military community.* Washington, DC, Headquarters, Department of Defense. Available: https://download.militaryonesource.mil/12038/MOS/Reports/2018-demographics-report.pdf [March 2020]

Department of Defense. (2018b). *Summary of the 2018 national defense strategy of the United States of America: Sharpening the American Military's competitive edge*. Washington DC: U.S. Government Printing Office. Available: https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf [March 2020]

Department of the Army. (2010). *Army health promotion, risk reduction, suicide prevention report 2010.* Washington, DC: U.S. Government Printing Office. Available: https://preventsuicide.army.mil/docs/Commanders%20Tool%20Kit/HPRRSP_Report_2010_v00.pdf [March 2020]

Department of the Army. (2014). *Army Regulation (AR) 350-53*, *Comprehensive soldier and family fitness.* Washington, DC: U.S. Government Printing Office. Available: https://readyandresilient.army.mil/CSF2/supportdocs/r350_53.pdf [March 2020]

Eidelson, R., Pilisuk, M., & Soldz, S. (2011). The dark side of Comprehensive Soldier Fitness. *American Psychologist*, *66*, 643–644. https://doi.org/10.1037/a0025272

Enders, C. K. (2010). *Applied missing data analysis*. New York: Guilford Press.

Fetzer Institute. (1999). *Multidimensional measurement of religiousness, spirituality for use in health research: A report of the National Working Group*, Kalamazoo, MI: Author.

Finch, W. H., Bolin, J. E., & Kelley, K. (2019). *Multilevel modeling using R* (2nd ed.). CRC Press.

Gregorich, S. E. (2006). Do self-report instruments allow meaningful comparisons across diverse population groups? Testing measurement invariance using the confirmatory factor analysis framework. *Medical Care*, *44*, S78–S94. https://doi.org/10.1097/01.mlr.0000245454.12228.8f

Harrison, X. A., Donaldson, L., Correa-Cano, M. E., Evans, J., Fisher, D. N., Goodwin, C. E. D., Robinson, B. S., Hodgson, D. J., & Inger, R. (2018). A brief introduction to mixed effects modeling and multi-model inference in ecology. *PeerJ*. https://doi.org/10.7717/peerj.4794

Hu, L.-t., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*, 1–55. https://doi.org/10.1080/10705519909540118

Ibáñez, M. I., Viruela, A. M., Mezquita, L., Moya, J., Villa, H., Camacho, L., & Ortet, G. (2016). An investigation of five types of personality trait continuity: A two-wave longitudinal study of Spanish adolescents from age 12 to age 15. *Frontiers in Psychology*, *7*, 1–7. https://doi.org/10.3389/fpsyg.2016.00512

Isiordia, M., & Ferrer, E. (2016). Curve of factors model: A latent growth modeling approach for educational research. *Educational and Psychological Measurement*, *78*, 203–231. https://doi.org/10.1177/0013164416677143

Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. J*ournal of Consulting and Clinical Psychology*, *59*, 12–19. https://doi.org/10.1037/0022-006X.59.1.12

Jöreskog, K. G., & Sörbom, D. (1986). *LISREL VI: Analysis of linear structural relationships by maximum likelihood, instrumental variables, and least squares methods*. Mooresville, IN: Scientific Software.

Knapp, D., Asch, B. J., DeMartini, C., Ruder, T., & Hanley, J. M. (2018). Using the Person-Event Data Environment for military personnel research in the Department of Defense: An evaluation of capability and potential uses. Santa Monica, CA: RAND Corporation, Available: https://www.rand.org/pubs/research_reports/RR2302.html [March 2020]

Kroenke, K., Spitzer, R. L., & Williams, J. B. (2001). The PHQ-9: Validity of a brief depression severity measure. *Journal of General Internal Medicine, 16,* 606–613. https://doi.org/10.1046/j.1525-1497.2001.016009606.x

Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for *t*-tests and ANOVAs. *Frontiers in Psychology*, *4*, 1–12. https://doi.org/10.3389/fpsyg.2013.0086

Larsen, R. (2011). Missing data imputation versus full information maximum likelihood with second-level dependencies. *Structural Equation Modeling: A Multidisciplinary Journal*, *18*, 649–662. https://doi.org/10.1080/10705511.2011.607721

Lester, P. B., Harms, P. D., Bulling, D. J., Herian, M. N., Spain, S. M. (2011). Evaluation of the relationships between reported resilience and outcomes – Report #1: Negative outcomes (suicide, drug use, & violent crimes). (DAMO-CSF Technical Report). Arlington, VA: Department of Army Military Operations. DTIC#ADA538618

Lester, P. B., Harms, P. D., Herian, M. N., & Sowden, W. J. (2015). A force for change: Chris Peterson and the US Army's Global Assessment Tool. *The Journal of Positive Psychology*, *10*, 7–16. https://doi.org/10.1080/17439760.2014.927904

Little, T. D. (2013). *Longitudinal structural equation modeling*. New York, NY: Guilford press.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment, 22,* 471–491. https://doi.org/10.1037/a0019227

Martin, M. M., & Rubin, R. B. (1995). A new measure of cognitive flexibility. *Psychological Reports, 76,* 623–626. http://dx.doi.org/10.2466/pr0.1995.76.2.623

Mayer, R. C. & Davis, J. H. (1999). The effect of the performance appraisal system on trust for management: A field quasi-experiment. *Journal of Applied Psychology*, *84*, 123–136. https://doi.org/10.1037/0021-9010.84.1.123

Mayer, R. C., Davis, J. H., & Schoorman, F. D. (1995). An integrative model of organizational trust. *The Academy of Management Review, 20,* 709–734. https://doi.org/10.2307/258792

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin*, *111*, 361–365. https://doi.org/10.1037/0033-2909.111.2.361

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*, 412–433. https://doi.org/10.1037/met0000144

Murphy, K. R., & Davidshofer, C. O. (1988). *Psychological testing: Principles and applications*. Prentice-Hall.

National Academies of Sciences, Engineering, and Medicine. (2017). *Strengthening Data Science Methods for Department of Defense Personnel and Readiness Missions*. National Academies Press.

National Academies of Sciences, Engineering, and Medicine. (2019). *A Decadal Survey of the Social and Behavioral Sciences: A Research Agenda for Advancing Intelligence Analysis*. National Academies Press.

Peterson, C. (2007). *Brief Strengths Test*. Cincinnati: VIS Institute.

Peterson, C., Bishop, M. P., Fletcher, C. W., Kaplan, M. R., Yesko, E. S., Moon, C. H., Smith, J. S., Michaels, C. E., & Michaels, A. J. (2001). Explanatory style as a risk factor for traumatic mishaps. *Cognitive Therapy and Research*, *25*, 633–649. https://doi.org/10.1023/A:1012945120821

Peterson, P., Park, N., & Castro, C. A. (2011). Assessment for the U.S. Army Comprehensive Soldier Fitness Program: The Global Assessment Tool. *American Psychologist*, *66*, 10–18. https://doi.org/10.1037/a0021658

Peterson, C., Park, N., & Seligman, M. E. P. (2005). Orientations to happiness and life satisfaction: The full life versus the empty life. *Journal of Happiness Studies*, *6*, 25–41. https://doi.org/10.1007/s10902-004-1278-z

Peterson, C., & Seligman, M. E. P. (2004). *Character Strengths and Virtues: A Handbook and Classification*. New York: Oxford University Press/Washington, DC: American Psychological Association.

Peterson, C., Semmel, A., Von Baeyer, C., Abramson, L. Y., Metalsky, G. I., & Seligman, M. E. (1982). The Attributional Style Questionnaire. *Cognitive Therapy and Research, 6,* 287–299. https://doi.org/10.1007/BF01173577

Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, *41*, 71–90. https://doi.org/10.1016/j.dr.2016.06.004

Roberts, B. W., Wood, D., & Caspi, A. (2008). The development of personality traits in adulthood. In O. P. John, R. W. Robins, and L. A. Pervin (Eds.), *Handbook of Personality: Theory and Research* (Vol. 3, pp. 375–398). The Guildford Press.

Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, *69*, 617–640. https://doi.org/10.1111/1467-6494.694157

Rosseel, Y. (2014). *Longitudinal structural equation modeling* [Course lecture slides]. Retrieved from http://www.personality-project.org/revelle/syllabi/454/rosseel_sem_longitudinal.pdf [March 2020]

Rosseel, Y. (2020). *Structural equation modeling with lavaan* [Course lecture slides]. Retrieved from https://users.ugent.be/~yrosseel/lavaan/gent2020/lavaan_twodays_gent2020.pdf [March 2020]

Russell, D., Peplau, L. A., & Ferguson, M. L. (1978). Developing a measure of loneliness. *Journal of Personality Assessment, 42,* 290–294. https://doi.org/10.1207/s15327752jpa4203_11

Scheier, M. F., & Carver, C. S. (1985). Optimism, coping, and health: Assessment and implications of generalized outcome expectancies. *Health Psychology, 4,* 219–247. https://doi.org/10.1037/0278-6133.4.3.219

Scheier, M. F., Carver, C. S., & Bridges, M. W. (1994). Distinguishing optimism from neuroticism (and trait anxiety, self-mastery, and self-esteem): A reevaluation of the Life Orientation Test. *Journal of Personality and Social Psychology, 67,* 1063–1078. https://doi.org/10.1037/0022-3514.67.6.1063

Seligman, M. E. P., & Csikszentmihalyi, M. (2000). Positive psychology: An introduction. *American Psychologist, 55,* 5–14. http://dx.doi.org/10.1037/0003-066X.55.1.5

Shingleton, J., Buttrey, S., Orendorff, F., & Wright, E. (2016). Global Assessment Tool (GAT) trend analysis. (TRAC-M-TR-18-012). Monterey, CA: TRADOC Analysis Center. DTIC# AD1050314

Spitzer, R. L., Kroenke, K., & Williams, J. B. (1999). Validation and utility of a self-report version of PRIME-MD: The PHQ Primary Care Study. Primary Care Evaluation of Mental Disorders. Patient Health Questionnaire. *JAMA: The Journal of the American Medical Association, 282,* 1737–1744. https://doi.org/10.1001/jama.282.18.1737

Steenkamp, M. M., Nash, W. P., & Litz, B. T. (2013). Post-traumatic stress disorder: Review of the Comprehensive Soldier Fitness program. *American Journal of Preventive Medicine, 44,* 507–512. http://dx.doi.org/10.1016/j.amepre.2013.01.013

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25,* 173–180. https://doi.org/10.1207/s15327906mbr2502_4

Steiger, J. H. (2004). Beyond the *F* test: Effect size confidence intervals and tests of close fit in the analysis of variance and contrast analysis. *Psychological Methods*, *9*, 164-182. https://doi.org/10.1037/1082-989X.9.2.164

Steiger, J. H., & Fouladi, R. T. (1997). Noncentrality interval estimation and the evaluation of statistical models. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests?* (pp. 221–257). Mahwah, NJ: Erlbaum.

Steyer, R., Mayer, A., Geiser, C., & Cole, D. A. (2015). A theory of states and traits--revised. *Annual Review of Clinical Psychology*, *11*, 71–98. https://doi.org/10.1146/annurev-clinpsy-032813-153719

Sweeney, P. J., Thompson, V., & Blanton, H. (2009). Trust and influence in combat: An interdependence model. *Journal of Applied Social Psychology, 39,* 235–264. https://doi.org/10.1111/j.1559-1816.2008.00437.x

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38,* 1–10. https://doi.org/10.1007/BF02291170

Vandenberg, R. J. (2002). Toward a further understanding of and improvement in measurement invariance methods and procedures. *Organizational Research Methods*, *5*, 139–158. https://doi.org/10.1177/1094428102005002001

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*, 4–70. https://doi.org/10.1177/109442810031002

Vargha, A., & Delaney, H. D. (2000). A critique and improvement of the CL common language effect size statistics of McGraw and Wong. *Journal of Educational and Behavioral Statistics*, *25*, 101–132. https://doi.org/10.3102/10769986025002101

Vie, L. L., Scheier, L. M., Lester, P. B., Ho, T. E., Labrthe, D. R., Seligman, M. E. P. (2015). The U.S. Army Person-Event Data Environment: A military-civilian big data enterprise. *Big Data*, *3*, 1–13. https://doi.org/10.1089/big.2014.0055

Vie, L. L., Scheier, L. M., Lester, P. B., Seligman, M. E. P. (2016). Initial validation of the U.S. Army Global Assessment Tool. *Military Psychology*, *28*, 468–487. https://doi.org/10.1037/mil0000141

Voelkel, M. C. (2007). Latent growth curve modeling as an integrative approach to the analysis of change. *Psychology Sciences*, *49*, 375–414.

Watson, D., & Clark, L. A. (1999). *The PANAS-X: Manual for the Positive and Negative Affect Schedule-Expanded Form*. Iowa City, IA: University of Iowa. https://doi.org/10.17077/48vt-m4t2

Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, *54*, 1063–1070. https://doi.org/10.1037//0022-3514.54.6.1063

Watson, D., & Walker, L. M. (1996). The long-term stability and predictive validity of trait measures of affect. *Journal of Personality and Social Psychology, 70*, 567–577. https://doi.org/10.1037/0022-3514.70.3.567

Widaman, K. F., Ferrer, E., & Conger, R. D. (2010). Factorial invariance within longitudinal structural equation models: Measuring the same construct across time. *Child Development Perspectives*, *4*, 10–18. https://doi.org/10.1111/j.1750-8606.2009.00110.x

Wrzesniewski, A., McCauley, C., Rozin, P., & Schwartz, B. (1997). Jobs, careers, and callings: People's relations to their work. *Journal of Research in Personality, 31,* 21–33. https://doi.org/10.1006/jrpe.1997.2162

Zuckerman, M. (1983). The distinction between trait and state scales is not arbitrary: Comment on Allen and Potkay's "On the arbitrary distinction between traits and states." *Journal of Personality and Social Psychology*, *5*, 1083–1086. https://doi.org/10.1037/0022-3514.44.5.1083

Appendix A: Detailed Description of Measures within the GAT

Description of GAT Measures

Adaptability

Three items were specifically developed for the GAT to assess adaptability. Adaptability items were designed to measure one's ability to change course or be cognitively flexible (Martin & Rubin, 1995). Using the stem "How well do these statements describe you", example items include "I am good at changing myself to adjust to changes in my life" or "I can usually fit myself into any situation." Respondents evaluated items using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Active Coping

Five items were adapted from the Brief COPE to assess active (or problem-focused) coping. Active coping items were designed to measure coping strategies that involve planning, or taking action (Carver, 1997; Carver et al., 1989). Using the stem "How well do these statements describe you", example items include "When something stresses me out, I try to solve the problem" or "I control my emotions by changing how I think about things". Respondents evaluated items using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).[1]

Passive Coping

Three items were adapted from the Brief COPE to assess passive (or emotion-focused) coping. Passive coping items were designed to measure coping strategies that involve venting, or disengagement (Carver, 1997; Carver et al., 1989). Using the stem "How well do these statements describe you", example items include "I usually keep my emotions to myself" (reverse-coded) or "When something stresses me out, I try to avoid it or not think about it" (reverse-coded). Respondents evaluated items using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Character

Twenty-four items were adapted from the Character Strengths Test, an abridged version of the VIA-IS (Peterson, 2007; Peterson & Seligman, 2004) to assess character strengths across different domains. Using the stem "Think about how you have acted in actual situations during the past four weeks", character items were designed to measure the character strengths of individuals across six character virtues. These virtues included knowledge (e.g., "critical thinking, open-mindedness, or good judgment"), courage (e.g., "persistence"), humanity (e.g., "kindness or generosity to others"), justice (e.g., "fairness"), temperance (e.g., "modesty or humility"), and transcendence (e.g., "gratitude"). Respondents evaluated items using an 11-point Likert scale (0 = *Never*, 10 = *Always*). In Study 2, six items were dropped for the GAT 2.0.

[1] Generally, items were coded such that higher numbers equaled more of the construct being measured. However, in the codebooks associated with the data, the suggested scaling is that higher numbers equal more desirable resilience traits.

Catastrophizing

Seven items were adapted from the Attributional Style Questionnaire (Peterson et al., 1982; Peterson et al., 2001) to assess catastrophic thinking. Catastrophizing items were designed to measure the explanatory style (i.e., stable, global, internal attributions) individuals form in response to negative events. Using the stem "Answer in terms of how you usually think", example items include "When bad things happen to me, I expect more bad things to happen" or "When I fail at something, I give up all hope". Respondents evaluated items using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*). In Study 2, four items were dropped for the GAT 2.0.

Depression

Eight items were adapted from the Patient Health Questionnaire (PHQ; Kroenke et al., 2001; Spitzer et al., 1999) and two items were generated for the military context to assess levels of depression. Depression items were designed to measure the degree to which respondents experienced depressive symptoms (e.g., feeling hopeless, tired, or down). Using the stem "In the past four weeks, how often have you been bothered by any of the following problems", example items from the PHQ include "Feeling down, depressed, or hopeless" or "Little interest or pleasure in doing things" and "Feeling very frustrated" or "Feeling very angry" from the military-generated items. Respondents evaluated items using a 5-point Likert scale (1 = *Not at all*, 5 = *Every day*).

Optimism

Four items were used from the Life Orientation Test (Scheier & Carver, 1985; Scheier et al., 1994) to assess optimism. Optimism items were designed to measure generalized optimism as manifest by the general expectation for positive future events or outcomes. Using the stem "Answer according to your own feelings, rather than how you think most people would answer", example items include "In uncertain times, I usually expect the best" or "I rarely count on good things happening to me" (reverse-coded). Respondents evaluated items using a 5-point Likert scale (1 = *Strongly disagree*, 5 = *Strongly agree*).

Positive Affect

Ten items were adapted from the Positive and Negative Affect Schedule Expanded Form (PANAS-X; Watson & Clark, 1999; Watson et al., 1988) to assess feelings of positive affect. Positive affect items were designed to measure subjective feelings of positive affect across general and specific dimensions. Using the stem "How often have you felt this way during the past four weeks", example items include those in a general dimension of positive affect (e.g., "Inspired"), and those across more specific dimensions of positive affect like self-assurance (i.e., "Proud"), joviality (e.g., "Joyful"), and serenity (e.g., "Peaceful/calm"). Respondents evaluated items using a 5-point Likert scale (1 = *Never*, 5 = *Most of the time*). In Study 2, one item was dropped for the GAT 2.0.

Negative Affect

Eleven items were adapted from the PANAS-X (Watson & Clark, 1999; Watson et al., 1988) to assess feelings of negative affect. Negative affect items were designed to measure subjective feelings of negative affect across general and specific dimensions. Using the stem "How often have you felt this way during the past four weeks", example items include those in a general dimension of negative affect (e.g., "Upset"), and those across more specific dimensions of negative affect like fear (i.e., "Scared/fearful"), hostility (e.g., "Hostile"), guilt (e.g., "Ashamed"), and sadness (e.g., "Sad"). Respondents evaluated items using a 5-point Likert scale (1 = *Never*, 5 = *Most of the time*). In Study 2, two items were dropped for the GAT 2.0.

Loneliness

Three items were adapted from the UCLA Loneliness Scale (Russell et al., 1978) to assess loneliness. Loneliness items were designed to measure subjective feelings of being alone or being set apart from groups. Using the stem "Please be honest as possible", example items include "How often do you feel left out" or "How often do you feel close to people" (reverse-coded). Respondents evaluated items using a 5-point Likert scale (1 = *Never*, 5 = *Most of the time*).

Organizational Trust

Five items were adapted from prior research on organizational trust (Mayer & Davis, 1999; Mayer et al., 1995; Sweeney et al., 2009) to assess trust in organizations. Organizational trust items were designed to measure three facets of organizational trust at the level of the unit. Using the stem "Please indicate how strongly you agree or disagree with each of the following statement", example items include those pertaining to facets of ability (e.g., "My immediate supervisor has much knowledge about the work that needs to be done"), benevolence (i.e., "I trust my fellow Soldiers in my unit to look out for my welfare and safety"), and integrity (e.g., "Overall, I trust my immediate supervisor"). Respondents evaluated items using a 5-point Likert scale (1 = *Strongly disagree*, 5 = *Strongly agree*). In Study 2, one item was dropped for the GAT 2.0.

Work Engagement

Four items were adapted from the Work as a Calling Scale (Wrzesniewski et al., 1997) and the Orientations to Happiness Scale (Peterson et al., 2005) to assess engagement in the workplace. Work engagement items were designed to measure aspects of work satisfaction, fulfillment, and commitment. Using the stem "How well do these statements describe your feelings about your job", example items include "My work is one of the most important things in my life" or "I am committed to my job." Respondents evaluated items using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Life Meaning

Five items were adapted from the Brief Multidimensional Measure of Religiousness/Spirituality (Fetzer Institute, 1999) and were crafted to be similar to the Purpose in Life Scale (Crumbaugh, 1968) to assess life meaning. Although originally framed as a measure of spiritual fitness or spirituality (Lester et al., 2011), life meaning items more accurately assess an individual's sense of purpose and meaning in the world and is not to be confused with measures of religiosity or religious belief (Peterson et al., 2011). Using the stem "Answer in terms of whether the statement describes how you actually live your life", example items include some worded generally "My life has meaning" or "I believe there is a purpose for my life" and others worded for a military-specific context (e.g., "The job I am doing in the military has enduring meaning"). Respondents evaluated items using a 5-point Likert scale (1 = *Not like me at all*, 5 = *Very much like me*).

Measures Not Examined

Following the guidance of prior work (Vie et al., 2016), three measures were not examined in our study. Family satisfaction and family support measures were excluded due to missing data for respondents who reported having no family or romantic relationships. In addition, we excluded a measure of friendships due to the measure's different response formats using five dichotomous (yes/no) items along with one ordered categorical item. In Study 2, we also excluded the physical activity and nutrition measures that were created for GAT 2.0. Examining these measures, they were composed mostly of non-interchangeable items that we felt were not related to an underlying psychological construct.

Appendix B: Detailed Description of Sample and Model-Testing Approaches

Sample for Study 1 (GAT 1.0) and Study 2 (GAT 2.0)

The two samples were created from GAT and administrative data stored in the Person-Event Data Environment (PDE) upon request and proper approvals. The GAT data spanned from 2009 to 2017 which covered the timeframe for both the GAT version 1.0 and version 2.0. Across both studies, respondents were active-duty Soldiers who were initially enlisted in the U.S. Army and provided consent to have their data used for study. Although we were unable to directly assess the GAT data of non-consenting respondents, we were able to analyze certain characteristics of these respondents and found that they were very similar to consenting respondents in terms of age, gender, race, and job specialty. For each version of the GAT, respondents were selected so that their accession date (start date) fell within the timeframe when the respective version of the GAT was administered (i.e., 2009–2014 for GAT 1.0 in Study 1 and 2012–2017 for GAT 2.0 in Study 2). In addition to accession date, we required that each respondent have completed data for at least two time occasions to be included in the sample. The maximum number of completed time occasions that was allowed for each respondent was five. Lastly, to ensure the GAT 2.0 sample in Study 2 was independent from the GAT 1.0 sample in Study 1, respondents were removed from the sample in Study 2 if data were present for them in Study 1.

The final sample for both studies was 95,277 respondents in Study 1 (GAT 1.0) and 57,771 respondents in Study 2 (GAT 2.0). The two samples were very similar in terms of age, gender, race, and job specialty. Analysis of missing data patterns indicated a predominant missing at random pattern such that missingness was attributed to either having two, three, four, or five completed time occasions [Study 1 *n*s: 95,277 (a single time occasion completed), 95,277 (two time occasions completed), 53,894 (three time occasions completed), 23,274 (four time occasions completed), 4,968 (five time occasions completed); Study 2 *n*s: 57,771 (a single time occasion completed), 57,771 (two time occasions completed), 19,194 (three time occasions completed), 4,722 (four time occasions completed), 679 (five time occasions completed)]. Again, across levels of completed time occasions, sub-samples were nearly identical in terms of age, gender, race, and job specialty. This pattern of missing at random data was planned in our cross-sequential longitudinal design due to the inclusion criteria for the samples. No other patterns of missing data were observed.

Model Effect Size Calculations across Models

Two measures of effect size were used across model-testing approaches, generalized η^2 and Cohen's ω. For RM-SEM, MI, and RM-CFA models, η^2 was calculated by transforming the χ^2 difference test statistic into an F statistic using the following formula: $F = (\Delta\chi^2 / \Delta df)$ (Y. Rosseel, personal communication, February 26, 2020). For RM-ANOVA, numerator and denominator degrees of freedom were readily available from models to calculate η^2, however, for RM-SEM, MI, and RM-CFA models, an estimate of denominator degrees of freedom (or error *df*) was calculated similar to RM-ANOVA by the following formula: $df_{den} = [(\text{no. of observations} - 1) - (N - 1) - (K - 1)]$. The Δdf from the χ^2 difference test was used as the numerator degree of freedom. Thus, η^2 calculations for these tests were a close approximation and did not

meaningfully change the pattern or conclusions of results when compared to simply using the sample size as the sole metric of denominator degree of freedom. Cohen's ω was calculated using the following formula: $\omega = \sqrt{\Delta\chi^2 / (N \times \Delta df)}$, where N is the sample size and $\Delta\chi^2$ and Δdf are the difference statistics between the two comparison models.

Repeated Measures Analysis of Variance (RM-ANOVA)

For each GAT measure, items were averaged together to form a scale composite variable. Moreover, five composites were created for each measure representing each of the five time occasions. To test if each measure was stable over time, we entered the five composites for each measure into a one-way repeated measures analysis of variance (RM-ANOVA) using the car package in R (v. 3.0.3). Covariance matrices were structured using compound symmetry such that all variances and covariances were set to be equal across time. From the ANOVA results, generalized η^2 was calculated along with 90% confidence intervals using a non-centrality interval estimation approach (Steiger, 2004; Steiger & Fouladi, 1997). Missing data were treated using a listwise deletion, limiting the sample to the number of respondents who completed all five time occasions. In addition, assumptions of normality and sphericity were tested; Q-Q plots and measures of skewness (values ranged from −1.23 to 1.96) and kurtosis (values ranged from −0.49 to 4.39) indicated that GAT measures were approximately normal and Greenhouse-Geisser epsilon (ε) values ranged from 0.89 to 0.99, indicating relatively small violations of the sphericity assumption (i.e., variance of differences between time occasions are equal).[2]

Thus, by design, the traditional ANOVA approach was a more conservative test of the null hypothesis that mean differences across five time occasions did not differ because the test focused only on those who had complete data for the full time course.[3]

Repeated Measures Structural Equation Model (RM-SEM)

Similar to the RM-ANOVA approach, we examined the data using a repeated measures approach within a structural equation modeling (RM-SEM) framework (Rosseel, 2014, 2020). Compared to the RM-ANOVA approach, analyzing the data using an SEM approach had the advantage of allowing for more flexible model selection and model-based approaches to missing data (Little, 2013). For each GAT measure, the same five scale composite variables (representing time occasions) were used as in the RM-ANOVA approach. To test if each measure was stable over time, we entered the five composites for each measure as observed scale indicators without any latent factors in a RM-SEM using the lavaan package in R (v. 0.6.4). Furthermore, two nested models were estimated for each measure with the first being a full model where the composite means were allowed to be freely estimated and a null (or restricted) model where the means were constrained to be equivalent across the time occasion composites.[4] To maintain parity with the covariance structure of the RM-ANOVA analyses, we set the structure of the RM-SEMs to have compound symmetry.[5] A maximum likelihood (ML) estimator was used in

[2] Corrected *p*-values did not meaningfully differ from the raw uncorrected *p*-values.
[3] For analogous results treating time as a continuous metric in a multi-level framework, see Tables I1 and I2.
[4] Further analyses estimating the random intercepts and random slopes of these models can be found in Tables J1 and J2.
[5] This assumption of covariance structure was tested compared to an unstructured method allowing variances and covariances to be freely estimated. However, this comparison was not meaningfully different in terms of model fits

analyses[6] and missing data were treated using a full-information maximum likelihood (FIML) approach which allowed parameter values and standard errors to be estimated in the presence of missing data (Enders, 2010; Larsen, 2011; Little, 2013). As a less-restrictive approach than listwise deletion used in the RM-ANOVA analyses, the FIML approach allowed for respondents with missing time occasions to be included in analyses, allowing for the full study sample size to be utilized. Model fit was evaluated using several indices including the comparative fit index (CFI; Bentler, 1990), the Tucker-Lewis fit index (TLI; Tucker & Lewis, 1973), the root mean square error of approximation with its 90% confidence intervals (RMSEA; Steiger, 1990), and the standardized root mean square residual (SRMR; Jöreskog & Sörbom, 1986). To compare models (full and null), a chi-square difference test was used along with measures of effect size (Cohen's ω, η^2) and changes between model fits for CFI, TLI, and RMSEA. In these comparison tests, if the null model (i.e., one that restricts the composite means to be equal) did not significantly differ from the full model (i.e., one that allows composite means to vary freely, without constraint) in a meaningful way, then it can be concluded that the given measure does not vary due to time occasions and is relatively stable.

Measurement Invariance (MI)

The third approach was a measurement invariance approach (MI). A MI approach was taken to test for different aspects of MI across time occasions and as an assurance that any potential differences between latent factors in the RM-CFA analyses (see below) could be clearly attributable to differences in latent means rather than other time-differing aspects of the factor structure like factor loadings or item intercepts (Gregorich, 2006; Putnick & Dornstein, 2016; Vandenberg, 2002; Vandenberg & Lance, 2000; Widaman et al., 2010). A series of longitudinal CFAs were conducted on each of the GAT measures (using lavaan package 0.6.4 in R) where each measured item was loaded onto a latent factor representing the unobserved measure or construct. Each item loading was freely estimated except for a single item which was set to a value of one. Five latent factors were modeled in each CFA such that the latent factors represented the five time occasions with their corresponding observed measurement items. Compound symmetry was assumed at the latent factor level and latent factors were allowed to covary between every time occasion. For analyses, a ML estimator was used along with FIML for handling missing data.

Across the five time occasions we tested for three types of MI: configural, weak (i.e., metric), and strong (i.e., scalar). Using the same model fit indices used in the RM-SEM analyses, the three types of MI were sequentially tested using chi-square difference tests along with measures of effect size (Cohen's ω, η^2). Each of these three MI types represents a nested hierarchy with increasingly restrictive equality constraints on the factor model (Vandenberg, 2002; Vandenberg & Lance, 2000). For *configural invariance*, the dimensional structure of the common factors and associated item indicators are set to be equivalent across time occasions. A well-fitting configural model would indicate that the factor structures across time occasions have

---

or patterns of results, thus we retained a compound symmetry structure. Moreover, this finding was true across the SEM-based approaches (RM-SEM, MI, and RM-CFA).

[6] We also tested models using a robust maximum likelihood (MLR) estimator, but results did not differ meaningfully compared to the ML approach. This result is not surprising given the large sample sizes and relatively small degrees of skewness in the data.

similar patterns of reflective item indicators. For *weak (metric) invariance* ($\lambda = \lambda$), the factor loadings were set to be equivalent across time occasions. A well-fitting model (weak invariance model fit or relative fit in comparison to the configural model using a chi-square difference test) would indicate that common factors have the same meaning across time occasions. For *strong (scalar) invariance* ($\tau = \tau$), the item intercepts are set to be equivalent across time occasions. A well-fitting model (strong invariance model fit or relative fit in comparison to the weak model using a chi-square difference test) would indicate a lack of systematic responses biases across the time occasions (e.g., at Time 2, all means for items increase by a constant of 1 compared to other time occasions). Of importance, if the assumption of strong invariance is supported then estimated factor means (see RM-CFA below) will be unbiased across time occasions allowing confidence in tests of equivalence of latent means (Putnick & Bornstein, 2016).

Repeated Measures Confirmatory Factor Analysis (RM-CFA)

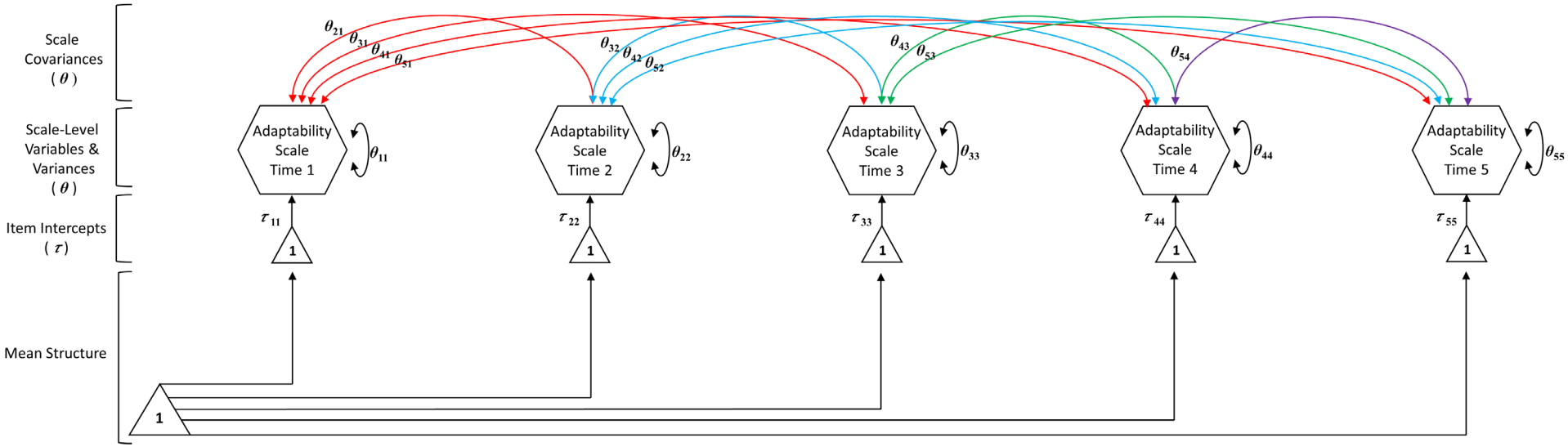
Similar to RM-SEM and MI, the final approach we took was a repeated confirmatory factor analysis (RM-CFA) approach also referred as a latent growth curve modeling (Voelkel, 2007) or structured means modeling (Breitsohl, 2019). This approach offered the most comprehensive accounting of measured item indicators and corresponding relationships between common factors for testing mean differences between factors across time occasions. For each GAT measure, a longitudinal CFA was conducted (using lavaan package 0.6.4 in R) where each measured item was freely loaded (one item set to value of one) onto a latent factor representing the unobserved measure or construct. To test if each measure was stable over time, the five latent factors were modeled in each RM-CFA such that the latent factors represented the five time occasions with their corresponding observed measurement items. Furthermore, two nested models were estimated for each measure with the first being a full model where the latent factor means were allowed to be freely estimated and a null (or restricted) model where the latent means were constrained to be equivalent across the time occasion factors. Compound symmetry was assumed at the latent factor level and latent factors were allowed to covary between every time occasion. For analyses, a ML estimator was used along with FIML for handling missing data. The same model fit indices used in the RM-SEM and MI analyses were used for RM-CFA. The fit of the two nested RM-CFA models were compared using chi-square difference tests along with measures of effect size (Cohen's ω, η^2). In these comparison tests, if the null model (i.e., one that restricts the latent means to be equal) does not significantly differ from the full model (i.e., one that allows latent means to vary freely, without constraint) in a meaningful way, then it can be concluded that the given measure does not vary due to time occasions and is relatively stable.

Appendix C: Example Conceptual Model Depictions

The following figures provide graphical depiction examples of the conceptual models analyzed across the various statistical techniques.

Figure C1

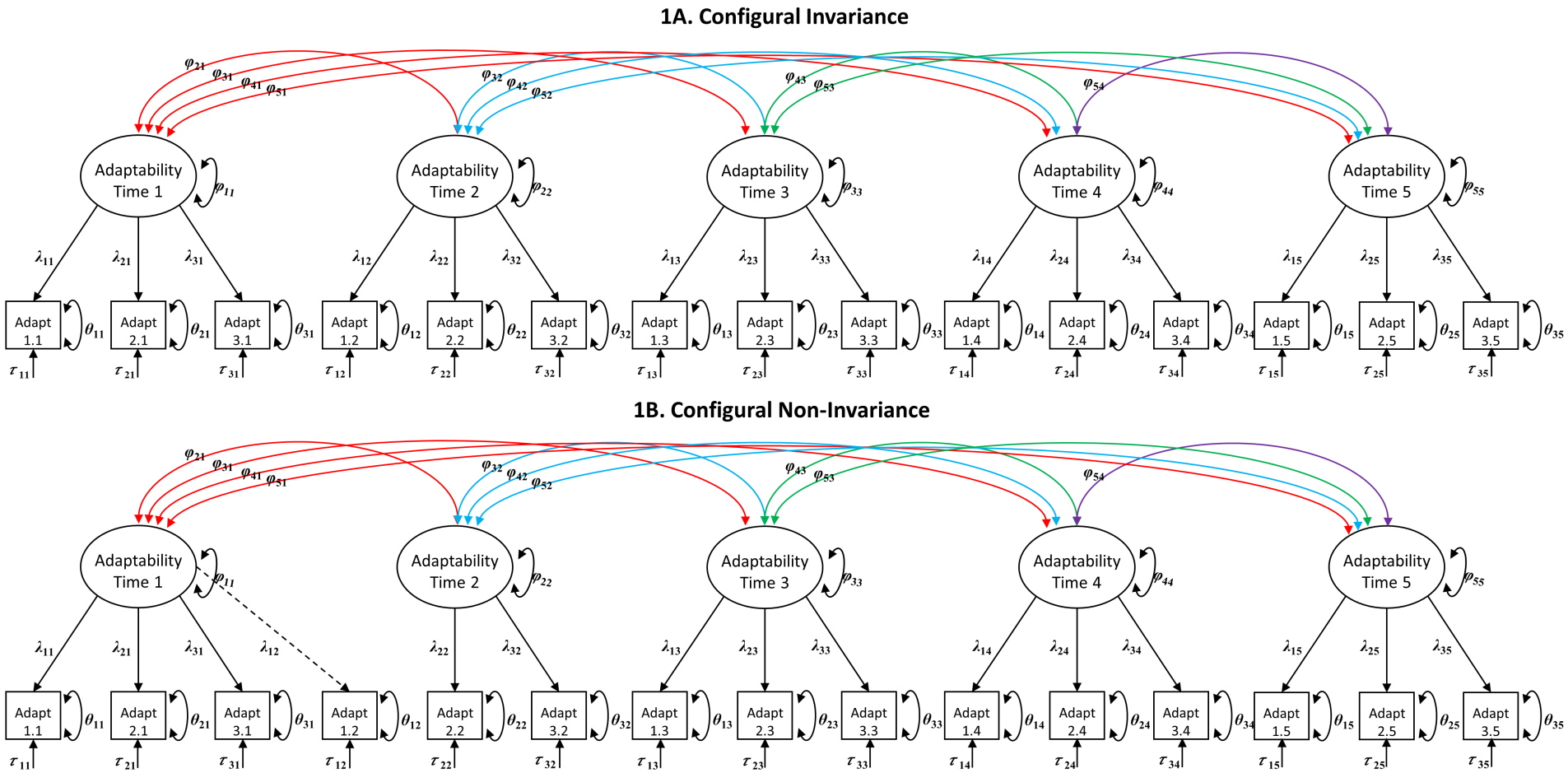
Repeated Measures Structural Equation Model (RM-SEM) Conceptual Model



Note. Hexagons represent scale composite (i.e., average of items of measure) across five time occasions. Triangles represent the mean structure on the scale intercepts across the five time occasions. Colored lines represent the covariances between scales at each of the five time occasions. For a more in-depth review, see Rosseel (2014, 2020).

Figure C2

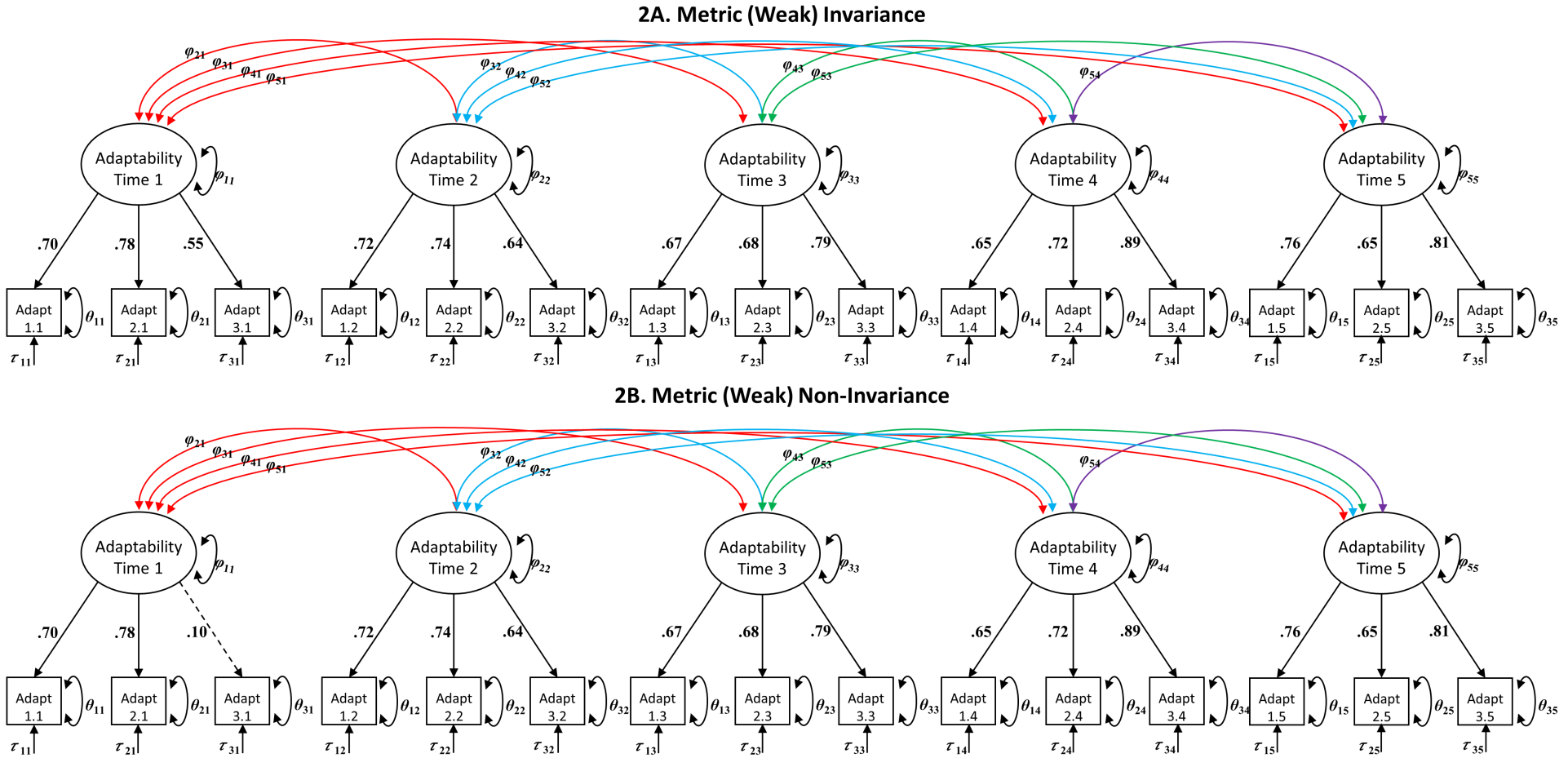
Configural Measurement Invariance Conceptual Model



Note. Circles represent latent factor variable reflecting upon three measured indicators across five time occasions. Panel 1A represents configural invariance (i.e., factor structure is constant across all time occasions). Panel 1B represents configural non-invariance (i.e., Item 1.2 is loading on the factor at Time 1 instead of Time 2 as expected). For a more in-depth review, see Putnick and Bornstein (2016) and Vandenberg and Lance (2000).

Figure C3

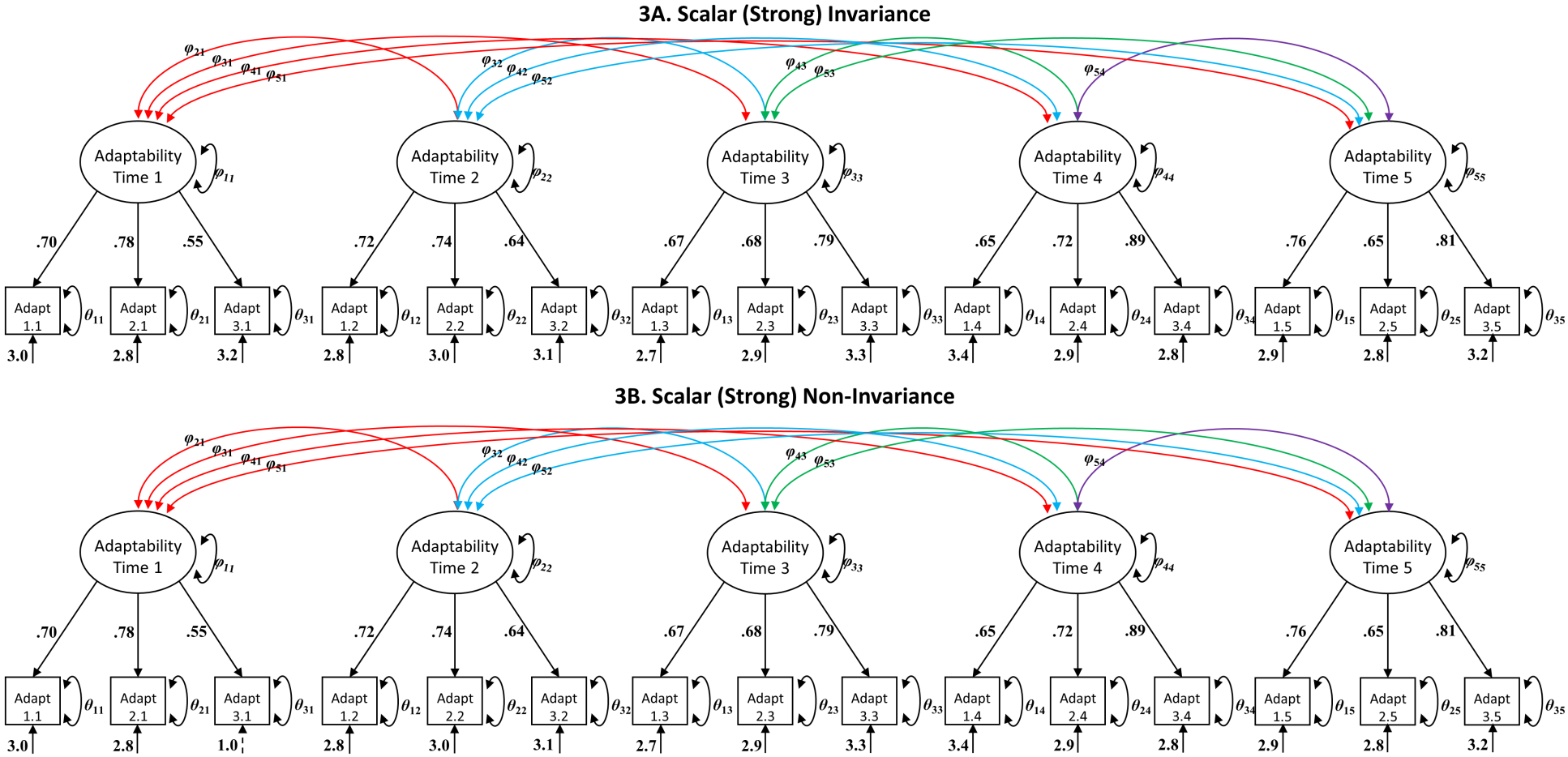
Metric (Weak) Measurement Invariance Conceptual Model



Note. Circles represent latent factor variable reflecting upon three measured indicators across five time occasions. Panel 2A represents metric (weak) invariance (i.e., factor loadings are constant across all time occasions). Panel 2B represents metric non-invariance [i.e., Item 3.1 is loading to a different degree (.10) compared to the other time occasions, e.g., 3.2 (.64), 3.3 (.79), 3.4 (.89), 3.5 (.81)]. For a more in-depth review, see Putnick and Bornstein (2016) and Vandenberg and Lance (2000).

Figure C4

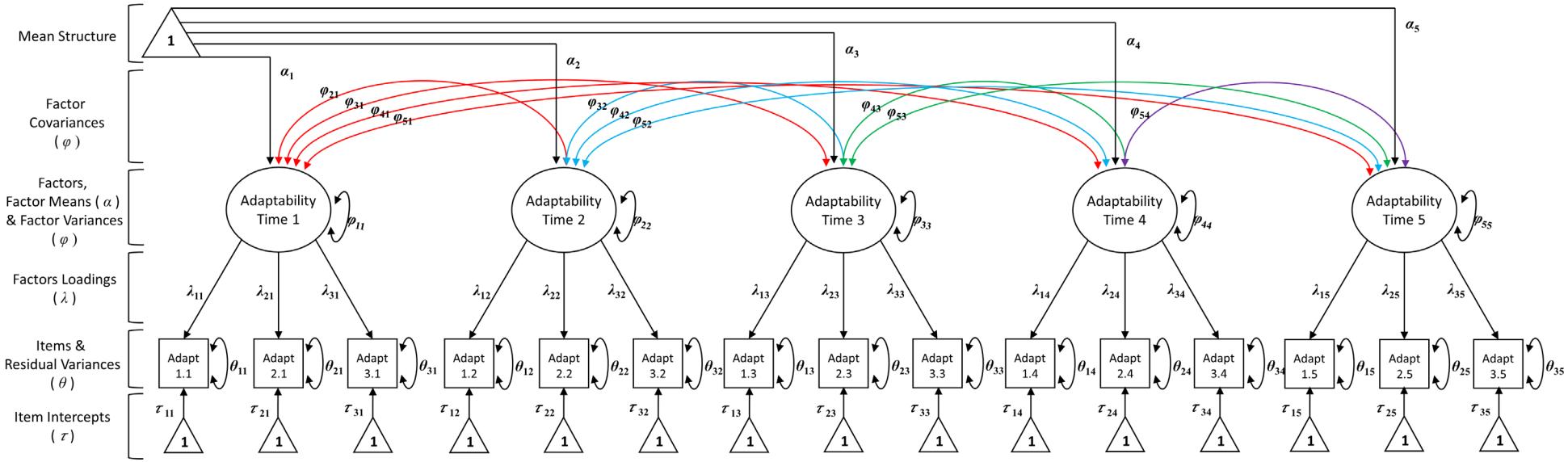
Scalar (Strong) Measurement Invariance Conceptual Model



Note. Circles represent latent factor variable reflecting upon three measured indicators across five time occasions. Panel 3A represents scalar (strong) invariance (i.e., item intercepts are constant across all time occasions). Panel 3B represents scalar non-invariance [i.e., Item 3.1's intercept varies to a different degree (1.0) compared to the other time occasions, e.g., 3.2 (3.1), 3.3 (3.3), 3.4 (2.8), 3.5 (3.2)]. For a more in-depth review, see Putnick and Bornstein (2016) and Vandenberg and Lance (2000).

Figure C5

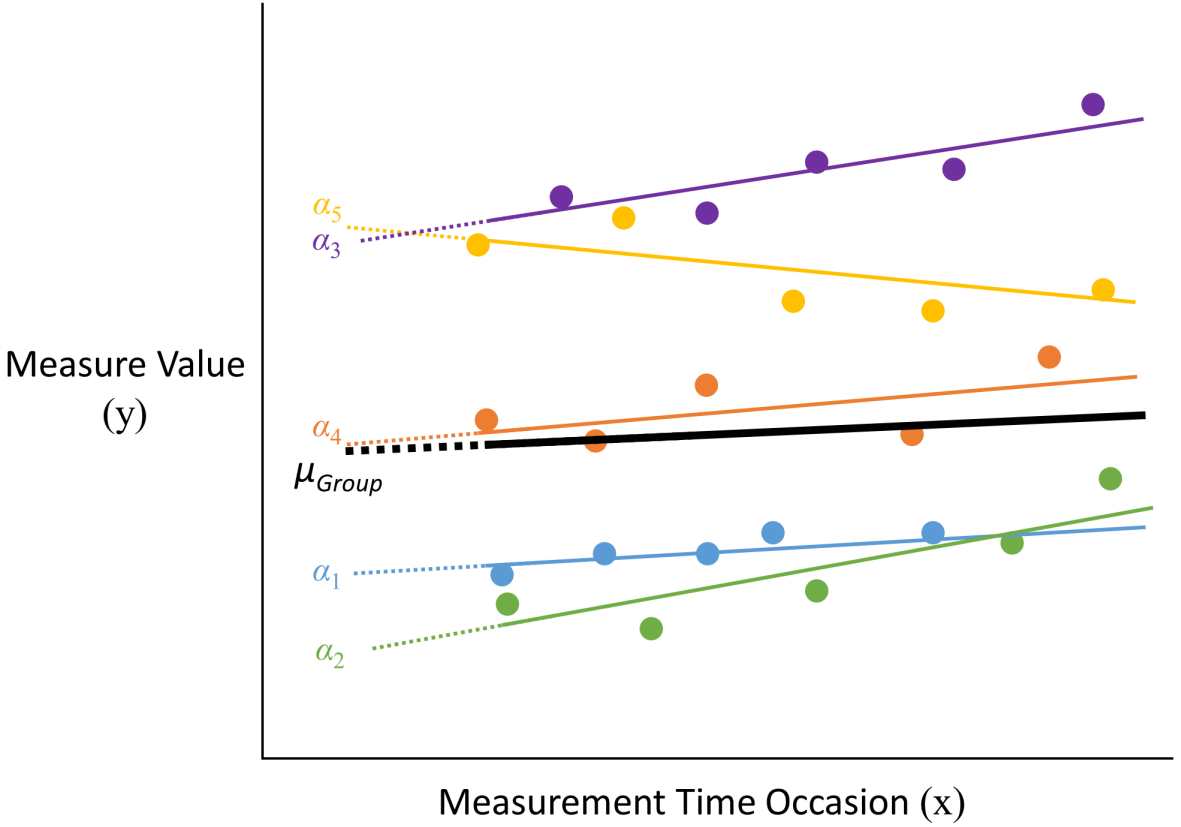
Repeated Measures Confirmatory Factor Analysis (RM-CFA) Conceptual Model



Note. Circles represent latent factor variable reflecting upon three measured indicators across five time occasions. Triangles represent the mean structure on the item intercepts and mean structure of the latent factors across the five time occasions. Colored lines represent the covariances between scales at each of the five time occasions. For a more in-depth review, see Isiordia & Ferrer (2016) and Rosseel (2014, 2020).

Figure C6

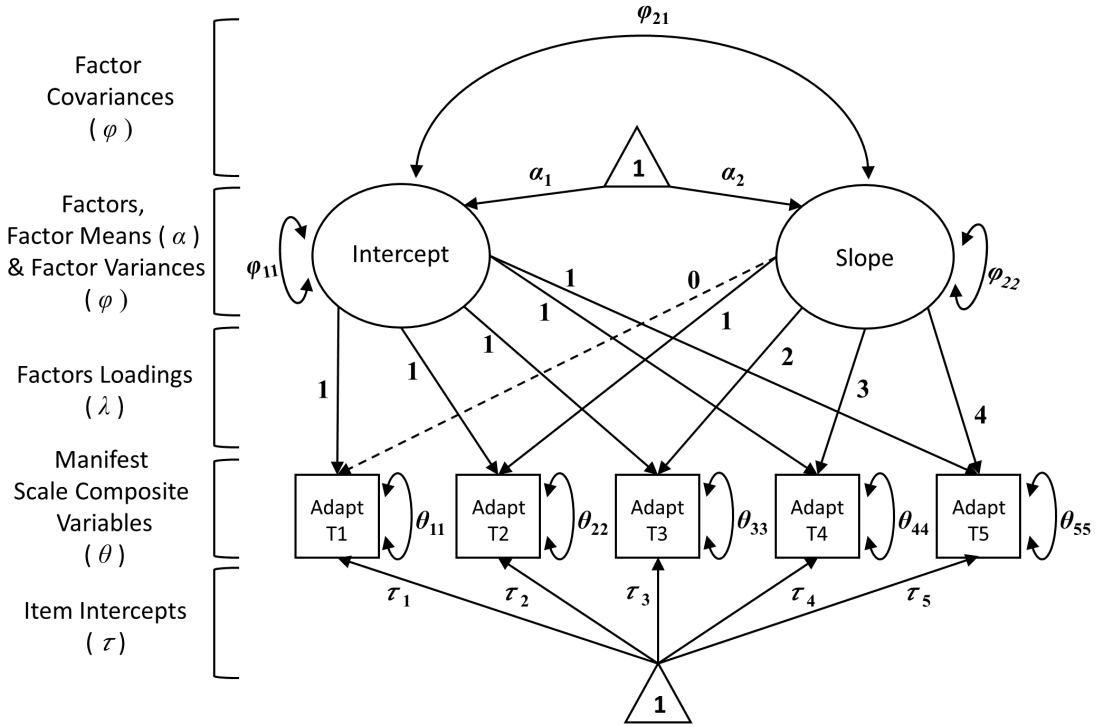
Repeated Measures Multi-Level Model (RM-MLM) Conceptual Model with Random Intercepts, Random Slopes, and Fixed Effect



Note. The conceptual depiction above represents a repeated multilevel model with respondents having nested repeated occasions of measurement (y) for a composite scale variable (e.g., Adaptability). The dots represent repeated measurement occasions (x) for each respondent accounting for the gaps of time in between each measurement (e.g., in months). The colored lines represent the random slopes for each respondent. The colored dotted lines represent the random intercepts for each respondent. The black line represents the mean slope for all the random slopes or simply, the fixed effect of time. For a more in-depth review, see Harrison et al. (2018) and Finch et al. (2019).

Figure C7

Growth Curve Structural Equation Model with Random Intercepts and Random Slopes



Note. Circles represent the latent intercept and latent slope for the latent variable (i.e., adaptability) across five time occasions. Triangles represent the mean structure on the item intercepts and mean structure of the latent factors across the five time occasions. For a more in-depth review, see Rosseel (2014, 2020).

Appendix D: Descriptive Statistics, Correlation Matrices, and Scale Reliabilities

Table D1

Scale-Level Pairwise Correlation Matrices, Descriptive Statistics, And Reliabilities GAT 1.0 Measures Across Time Occasions

		Time Occasion Bivariate Correlations					Descriptive Statistics					Reliabilities	
Measure		T1	T2	T3	T4	T5	# Items	*n*	*M*	95% CI_M	*SD*	ω_{Total}	95% CI ω_{Total}
Adaptability	T1	—					3	95,274	4.06	[4.06, 4.06]	0.74	.72	[.78, .73]
	T2	.42 (.58)	—				3	95,261	4.05	[4.04, 4.05]	0.77	.72	[.72, .73]
	T3	.38 (.52)	.48 (.66)	—			3	53,890	4.06	[4.06, 4.07]	0.78	.73	[.72, .74]
	T4	.35 (.48)	.42 (.58)	.50 (.68)	—		3	23,266	4.07	[4.06, 4.08]	0.78	.73	[.72, .74]
	T5	.35 (.48)	.41 (.57)	.46 (.63)	.52 (.71)	—	3	4,967	4.09	[4.07, 4.11]	0.77	.73	[.71, .74]
Active Coping	T1	—					5	95,272	3.85	[3.84, 3.85]	0.71	.75	[.75, .76]
	T2	.43 (.55)	—				5	95,261	3.82	[3.81, 3.82]	0.79	.83	[.82, .83]
	T3	.38 (.48)	.50 (.60)	—			5	53,890	3.82	[3.82, 3.83]	0.80	.84	[.84, .84]
	T4	.36 (.45)	.45 (.54)	.53 (.63)	—		5	23,266	3.84	[3.83, 3.85]	0.81	.85	[.85, .86]
	T5	.36 (.45)	.43 (.51)	.50 (.59)	.55 (.64)	—	5	4,967	3.87	[3.84, 3.89]	0.80	.86	[.85, .87]
Passive Coping	T1	—					3	95,272	2.60	[2.60, 2.61]	0.92	.70	[.70, .71]
	T2	.37 (.51)	—				3	95,261	2.68	[2.68, 2.69]	0.96	.74	[.74, .74]
	T3	.34 (.47)	.41 (.55)	—			3	53,890	2.74	[2.73, 2.75]	0.97	.75	[.74, .75]
	T4	.34 (.47)	.37 (.50)	.44 (.59)	—		3	23,266	2.77	[2.75, 2.78]	0.98	.75	[.75, .76]
	T5	.32 (.44)	.37 (.49)	.42 (.56)	.45 (.60)	—	3	4,967	2.80	[2.77, 2.82]	1.00	.76	[.74, .77]
Character	T1	—					24	95,273	7.98	[7.97, 7.99]	1.37	.95	[.95, .95]
	T2	.43 (.45)	—				24	95,258	7.97	[7.96, 7.98]	1.60	.97	[.97, .97]
	T3	.38 (.40)	.51 (.53)	—			24	53,884	7.96	[7.95, 7.98]	1.66	.97	[.97, .98]
	T4	.35 (.36)	.44 (.45)	.54 (.55)	—		24	23,267	7.99	[7.96, 8.01]	1.69	.98	[.98, .98]
	T5	.34 (.35)	.43 (.44)	.48 (.49)	.57 (.58)	—	24	4,966	8.02	[7.97, 8.06]	1.70	.98	[.98, .98]
Catastrophizing	T1	—					7	95,274	1.93	[1.93, 1.93]	0.73	.83	[.83, .84]
	T2	.39 (.46)	—				7	95,263	1.88	[1.87, 1.88]	0.80	.88	[.88, .88]
	T3	.37 (.43)	.45 (.51)	—			7	53,890	1.86	[1.85, 1.87]	0.81	.89	[.89, .89]
	T4	.35 (.41)	.41 (.46)	.49 (.55)	—		7	23,269	1.85	[1.83, 1.86]	0.82	.89	[.89, .90]
	T5	.33 (.38)	.40 (.45)	.45 (.50)	.49 (.55)	—	7	4,968	1.84	[1.81, 1.86]	0.84	.90	[.90, .91]
Depression	T1	—					10	95,273	1.60	[1.60, 1.61]	0.72	.91	[.91, .91]
	T2	.36 (.39)	—				10	95,267	1.65	[1.64, 1.65]	0.81	.94	[.93, .94]
	T3	.30 (.32)	.44 (.47)	—			10	53,893	1.64	[1.64, 1.65]	0.80	.94	[.93, .94]
	T4	.27 (.29)	.39 (.41)	.48 (.51)	—		10	23,270	1.64	[1.63, 1.65]	0.81	.94	[.94, .94]
	T5	.27 (.29)	.36 (.38)	.41 (.44)	.48 (.51)	—	10	4,967	1.63	[1.61, 1.66]	0.81	.94	[.94, .94]
Optimism	T1	—					4	95,273	3.71	[3.71, 3.72]	0.79	.71	[.71, .71]
	T2	.47 (.64)	—				4	95,263	3.68	[3.67, 3.68]	0.80	.76	[.76, .77]
	T3	.43 (.58)	.54 (.71)	—			4	53,890	3.70	[3.69, 3.70]	0.81	.77	[.77, .78]
	T4	.41 (.55)	.49 (.64)	.57 (.74)	—		4	23,269	3.73	[3.72, 3.74]	0.82	.78	[.77, .78]
	T5	.41 (.55)	.48 (.62)	.53 (.68)	.60 (.77)	—	4	4,968	3.74	[3.72, 3.77]	0.83	.78	[.77, .80]

Note. GAT = Global Assessment Tool. All bivariate correlations, *p*s < .001, correlations in parentheses represent disattenuated correlations accounting for measure reliability (see Murphy & Davidshofer, 1988); *n* = number of non-missing observations for a given measure and time occasion; CI = confidence interval; ω_{Total} = Omega Total.

Table D1 (continued)

Scale-Level Pairwise Correlation Matrices, Descriptive Statistics, and Reliabilities GAT 1.0 Measures Across Time Occasions

		Time Occasion Bivariate Correlations					Descriptive Statistics					Reliabilities	
Measure		T1	T2	T3	T4	T5	# Items	*n*	*M*	95% CI_M	*SD*	ω_{Total}	95% CI ω_{Total}
Positive Affect	T1	—					10	95,276	3.82	[3.81, 3.82]	0.77	.92	[.92, .92]
	T2	.45 (.48)	—				10	95,270	3.80	[3.79, 3.80]	0.84	.95	[.95, .95]
	T3	.40 (.43)	.53 (.56)	—			10	53,891	3.81	[3.80, 3.82]	0.84	.95	[.95, .95]
	T4	.37 (.40)	.47 (.49)	.56 (.59)	—		10	23,271	3.84	[3.83, 3.85]	0.85	.95	[.95, .95]
	T5	.38 (.40)	.46 (.48)	.51 (.53)	.59 (.62)	—	10	4,968	3.86	[3.84, 3.88]	0.87	.96	[.95, .96]
Negative Affect	T1	—					11	95,276	2.20	[2.20, 2.21]	0.68	.88	[.88, .88]
	T2	.40 (.45)	—				11	95,270	2.19	[2.18, 2.19]	0.73	.91	[.91, .91]
	T3	.35 (.39)	.47 (.52)	—			11	53,891	2.18	[2.17, 2.19]	0.73	.91	[.91, .91]
	T4	.32 (.36)	.42 (.46)	.51 (.56)	—		11	23,271	2.16	[2.15, 2.17]	0.73	.91	[.91, .91]
	T5	.32 (.36)	.40 (.44)	.42 (.46)	.51 (.56)	—	11	4,968	2.14	[2.11, 2.16]	0.74	.92	[.91, .92]
Loneliness	T1	—					3	95,273	2.21	[2.20, 2.21]	0.85	.80	[.80, .80]
	T2	.47 (.58)	—				3	95,263	2.21	[2.20, 2.21]	0.90	.83	[.82, .83]
	T3	.42 (.52)	.52 (.63)	—			3	53,890	2.20	[2.19, 2.20]	0.91	.83	[.83, .84]
	T4	.41 (.50)	.47 (.56)	.55 (.66)	—		3	23,269	2.20	[2.18, 2.21]	0.92	.84	[.84, .84]
	T5	.42 (.51)	.47 (.56)	.51 (.61)	.57 (.68)	—	3	4,968	2.20	[2.17, 2.22]	0.92	.84	[.83, .85]
Organizational Trust	T1	—					5	95,274	3.98	[3.97, 3.98]	0.75	.81	[.81, .82]
	T2	.31 (.37)	—				5	95,265	3.81	[3.80, 3.81]	0.90	.87	[.87, .88]
	T3	.24 (.28)	.37 (.42)	—			5	53,889	3.72	[3.72, 3.73]	0.93	.88	[.87, .88]
	T4	.22 (.26)	.31 (.35)	.40 (.45)	—		5	23,269	3.70	[3.69, 3.71]	0.94	.88	[.88, .89]
	T5	.20 (.24)	.29 (.33)	.33 (.37)	.43 (.49)	—	5	4,968	3.72	[3.70, 3.75]	0.94	.89	[.89, .90]
Work Engagement	T1	—					4	95,273	3.87	[3.86, 3.87]	0.88	.80	[.79, .80]
	T2	.39 (.48)	—				4	95,263	3.60	[3.59, 3.60]	1.01	.84	[.84, .84]
	T3	.31 (.38)	.50 (.60)	—			4	53,890	3.57	[3.56, 3.58]	1.01	.84	[.84, .85]
	T4	.28 (.34)	.42 (.50)	.52 (.62)	—		4	23,269	3.61	[3.59, 3.62]	1.00	.84	[.84, .85]
	T5	.29 (.35)	.39 (.46)	.48 (.57)	.57 (.67)	—	4	4,968	3.65	[3.62, 3.67]	1.00	.85	[.84, .86]
Life Meaning	T1	—					5	95,272	3.87	[3.86, 3.88]	0.86	.80	[.79, .80]
	T2	.50 (.61)	—				5	95,261	3.91	[3.91, 3.92]	0.92	.84	[.84, .84]
	T3	.40 (.49)	.52 (.62)	—			5	53,890	4.04	[4.03, 4.05]	0.88	.85	[.85, .85]
	T4	.37 (.44)	.46 (.54)	.56 (.65)	—		5	23,266	4.13	[4.12, 4.14]	0.86	.87	[.86, .87]
	T5	.36 (.43)	.44 (.51)	.52 (.60)	.58 (.66)	—	5	4,967	4.15	[4.13, 4.18]	0.86	.88	[.87, .88]

Note. GAT = Global Assessment Tool. All bivariate correlations, *p*s < .001, correlations in parentheses represent disattenuated correlations accounting for measurement error (see Murphy & Davidshofer, 1988); *n* = number of non-missing observations for a given measure and time occasion; CI = confidence interval; ω_{Total} = Omega Total.

Table D2

Scale-Level Pairwise Correlation Matrices, Descriptive Statistics, And Reliabilities GAT 2.0 Measures Across Time Occasions

		Time Occasion Bivariate Correlations					Descriptive Statistics					Reliabilities	
Measure		T1	T2	T3	T4	T5	# Items	*n*	*M*	95% CI_M	*SD*	ω_{Total}	95% CI ω_{Total}
Adaptability	T1	—					3	57,766	4.09	[4.09, 4.10]	0.73	0.69	[0.68, 0.69]
	T2	.45 (.65)	—				3	57,724	4.02	[4.02, 4.03]	0.77	0.69	[0.69, 0.70]
	T3	.42 (.60)	.50 (.72)	—			3	19,158	4.03	[4.02, 4.04]	0.77	0.70	[0.69, 0.71]
	T4	.40 (.57)	.49 (.70)	.56 (.79)	—		3	4,696	4.04	[4.02, 4.06]	0.78	0.71	[0.69, 0.72]
	T5	.45 (.63)	.48 (.67)	.51 (.71)	.56 (.77)	—	3	668	4.01	[3.95, 4.07]	0.80	0.74	[0.69, 0.78]
Active Coping	T1	—					5	57,765	3.94	[3.93, 3.94]	0.70	0.76	[0.76, 0.77]
	T2	.47 (.59)	—				5	57,724	3.88	[3.87, 3.88]	0.78	0.83	[0.83, 0.83]
	T3	.42 (.53)	.52 (.62)	—			5	19,158	3.87	[3.86, 3.88]	0.78	0.84	[0.83, 0.84]
	T4	.41 (.51)	.51 (.61)	.55 (.65)	—		5	4,696	3.89	[3.86, 3.91]	0.78	0.85	[0.84, 0.86]
	T5	.48 (.60)	.52 (.62)	.52 (.62)	.59 (.69)	—	5	668	3.86	[3.80, 3.92]	0.81	0.85	[0.83, 0.87]
Passive Coping	T1	—					3	57,765	2.50	[2.49, 2.51]	0.90	0.69	[0.68, 0.69]
	T2	.37 (.52)	—				3	57,724	2.52	[2.51, 2.52]	0.92	0.72	[0.72, 0.72]
	T3	.34 (.48)	.39 (.54)	—			3	19,158	2.56	[2.55, 2.58]	0.94	0.73	[0.72, 0.74]
	T4	.34 (.48)	.39 (.53)	.44 (.60)	—		3	4,696	2.61	[2.58, 2.64]	0.94	0.74	[0.72, 0.75]
	T5	.34 (.47)	.34 (.46)	.36 (.49)	.44 (.59)	—	3	668	2.59	[2.52, 2.66]	0.95	0.75	[0.71, 0.78]
Character	T1	—					18	57,764	8.18	[8.17, 8.19]	1.32	0.94	[0.94, 0.94]
	T2	.49 (.52)	—				18	57,724	8.08	[8.07, 8.09]	1.57	0.96	[0.96, 0.96]
	T3	.45 (.47)	.54 (.56)	—			18	19,158	8.08	[8.05, 8.10]	1.60	0.97	[0.97, 0.97]
	T4	.44 (.46)	.54 (.56)	.60 (.62)	—		18	4,696	8.11	[8.07, 8.16]	1.58	0.97	[0.97, 0.97]
	T5	.49 (.51)	.53 (.55)	.57 (.59)	.65 (.67)	—	18	668	8.02	[7.90, 8.15]	1.66	0.97	[0.97, 0.98]
Catastrophizing	T1	—					3	57,765	1.92	[1.91, 1.93]	0.87	0.73	[0.72, 0.73]
	T2	.40 (.53)	—				3	57,723	2.03	[2.03, 2.04]	0.96	0.79	[0.78, 0.79]
	T3	.36 (.47)	.43 (.54)	—			3	19,158	2.03	[2.01, 2.04]	0.96	0.80	[0.79, 0.81]
	T4	.36 (.47)	.40 (.50)	.45 (.56)	—		3	4,697	1.98	[1.95, 2.01]	0.95	0.81	[0.80, 0.82]
	T5	.41 (.53)	.46 (.58)	.48 (.60)	.52 (.64)	—	3	668	2.07	[1.99, 2.14]	1.00	0.81	[0.78, 0.84]
Depression	T1	—					10	57,764	1.51	[1.50, 1.51]	0.67	0.91	[0.91, 0.91]
	T2	.39 (.42)	—				10	57,721	1.60	[1.59, 1.61]	0.78	0.94	[0.93, 0.94]
	T3	.32 (.35)	.46 (.49)	—			10	19,157	1.62	[1.61, 1.63]	0.79	0.94	[0.93, 0.94]
	T4	.30 (.32)	.42 (.45)	.48 (.51)	—		10	4,697	1.60	[1.58, 1.63]	0.78	0.94	[0.93, 0.94]
	T5	.33 (.36)	.46 (.49)	.49 (.52)	.54 (.57)	—	10	668	1.62	[1.56, 1.68]	0.80	0.94	[0.93, 0.95]
Optimism	T1	—					4	57,491	3.71	[3.71, 3.72]	0.73	0.71	[0.71, 0.72]
	T2	.53 (.72)	—				4	35,912	3.66	[3.65, 3.67]	0.78	0.76	[0.75, 0.76]
	T3	.49 (.66)	.58 (.76)	—			4	9,101	3.67	[3.66, 3.69]	0.80	0.77	[0.76, 0.78]
	T4	.47 (.63)	.55 (.71)	.61 (.78)	—		4	1,235	3.64	[3.60, 3.69]	0.84	0.79	[0.76, 0.81]
	T5	.49 (.63)	.43 (.54)	.53 (.66)	.60 (.74)	—	4	147	3.73	[3.58, 3.88]	0.92	0.84	[0.77, 0.90]

Note. GAT = Global Assessment Tool. All bivariate correlations, *p*s < .001, correlations in parentheses represent disattenuated correlations accounting for measurement error (see Murphy & Davidshofer, 1988); *n* = number of non-missing observations for a given measure and time occasion; CI = confidence interval; ω_{Total} = Omega Total.

Table D2 (continued)

Scale-Level Pairwise Correlation Matrices, Descriptive Statistics, And Reliabilities GAT 2.0 Measures Across Time Occasions

		Time Occasion Bivariate Correlations					Descriptive Statistics					Reliabilities	
Measure		T1	T2	T3	T4	T5	# Items	*n*	*M*	95% CI_M	*SD*	ω_{Total}	95% CI ω_{Total}
Positive Affect	T1	—					9	57,489	3.94	[3.94, 3.95]	0.75	0.92	[0.92, 0.92]
	T2	.49 (.52)	—				9	35,908	3.87	[3.86, 3.88]	0.83	0.95	[0.94, 0.95]
	T3	.45 (.48)	.57 (.60)	—			9	9,100	3.85	[3.84, 3.87]	0.84	0.95	[0.95, 0.95]
	T4	.44 (.47)	.55 (.58)	.61 (.64)	—		9	1,235	3.82	[3.77, 3.87]	0.88	0.95	[0.95, 0.96]
	T5	.48 (.51)	.52 (.54)	.57 (.60)	.67 (.70)	—	9	147	3.92	[3.77, 4.07]	0.93	0.96	[0.95, 0.97]
Negative Affect	T1	—					9	57,489	2.02	[2.02, 2.03]	0.69	0.88	[0.87, 0.88]
	T2	.44 (.49)	—				9	35,908	2.04	[2.03, 2.05]	0.73	0.90	[0.90, 0.90]
	T3	.37 (.42)	.50 (.56)	—			9	9,100	2.05	[2.04, 2.07]	0.72	0.90	[0.89, 0.90]
	T4	.35 (.39)	.47 (.52)	.53 (.59)	—		9	1,235	2.06	[2.02, 2.10]	0.76	0.91	[0.90, 0.92]
	T5	.39 (.44)	.48 (.53)	.53 (.59)	.54 (.60)	—	9	147	2.10	[1.97, 2.23]	0.81	0.90	[0.87, 0.93]
Loneliness	T1	—					3	57,765	2.21	[2.21, 2.22]	0.82	0.79	[0.79, 0.79]
	T2	.50 (.63)	—				3	57,722	2.26	[2.26, 2.27]	0.88	0.80	[0.80, 0.81]
	T3	.46 (.58)	.56 (.70)	—			3	19,158	2.26	[2.24, 2.27]	0.88	0.81	[0.81, 0.82]
	T4	.44 (.54)	.52 (.64)	.58 (.71)	—		3	4,697	2.24	[2.22, 2.27]	0.90	0.83	[0.82, 0.84]
	T5	.45 (.55)	.45 (.54)	.54 (.65)	.60 (.71)	—	3	668	2.27	[2.20, 2.34]	0.93	0.86	[0.84, 0.88]
Organizational Trust	T1	—					4	57,766	4.10	[4.10, 4.11]	0.75	0.83	[0.83, 0.83]
	T2	.34 (.40)	—				4	57,724	3.90	[3.89, 3.91]	0.89	0.88	[0.88, 0.88]
	T3	.27 (.31)	.40 (.45)	—			4	19,158	3.83	[3.81, 3.84]	0.94	0.89	[0.88, 0.89]
	T4	.26 (.30)	.33 (.37)	.44 (.49)	—		4	4,697	3.83	[3.80, 3.86]	0.93	0.89	[0.88, 0.89]
	T5	.28 (.33)	.30 (.34)	.39 (.44)	.47 (.53)	—	4	668	3.81	[3.74, 3.89]	0.94	0.88	[0.86, 0.90]
Work Engagement	T1	—					4	57,766	3.96	[3.95, 3.96]	0.84	0.80	[0.80, 0.81]
	T2	.43 (.52)	—				4	57,724	3.67	[3.67, 3.68]	0.98	0.85	[0.85, 0.85]
	T3	.37 (.45)	.52 (.62)	—			4	19,158	3.63	[3.61, 3.64]	0.99	0.84	[0.84, 0.85]
	T4	.34 (.41)	.47 (.55)	.56 (.66)	—		4	4,697	3.64	[3.61, 3.67]	1.00	0.85	[0.85, 0.86]
	T5	.30 (.36)	.43 (.51)	.57 (.67)	.60 (.71)	—	4	668	3.67	[3.59, 3.74]	1.01	0.85	[0.83, 0.87]
Life Meaning	T1	—					5	57,765	4.32	[4.31, 4.32]	0.69	0.82	[0.81, 0.82]
	T2	.50 (.60)	—				5	57,724	4.19	[4.19, 4.20]	0.80	0.86	[0.86, 0.86]
	T3	.45 (.53)	.56 (.65)	—			5	19,158	4.17	[4.16, 4.19]	0.83	0.87	[0.86, 0.87]
	T4	.44 (.52)	.54 (.62)	.60 (.69)	—		5	4,697	4.19	[4.16, 4.21]	0.83	0.87	[0.86, 0.88]
	T5	.46 (.54)	.51 (.58)	.57 (.65)	.65 (.74)	—	5	668	4.18	[4.11, 4.24]	0.85	0.89	[0.88, 0.91]

Note. GAT = Global Assessment Tool. All bivariate correlations, *p*s < .001, correlations in parentheses represent disattenuated correlations accounting for measurement error (see Murphy & Davidshofer, 1988); *n* = number of non-missing observations for a given measure and time occasion; CI = confidence interval; ω_{Total} = Omega Total.

Appendix E: Repeated Measures Analysis of Variance (RM-ANOVA) Results

Table E1

Table of Repeated Measures ANOVA (RM-ANOVA) Tests Across Five Time Occasions for Each GAT 1.0 Measure

Measure	$\varepsilon_{G\text{-}G}$	F (df)	p	η^2	90% CI η^2
Adaptability	0.98	1.07 (4, 19,864)	.369	.000	[.000, .001]
Active Coping	0.99	4.64 (4, 19,864)	.001	.000	[.000, .001]
Passive Coping	0.98	42.12 (4, 19,864)	< .001*	.004	[.003, .006]
Character	0.96	3.37 (4, 19,860)	.010	.000	[.000, .001]
Catastrophizing	0.97	42.92 (4, 19,860)	< .001*	.004	[.003, .005]
Depression	0.98	6.50 (4, 19,864)	< .001*	.001	[.000, .001]
Optimism	0.97	8.09 (4, 19,860)	< .001*	.001	[.000, .001]
Positive Affect	0.95	3.60 (4, 19,868)	.006	.000	[.000, .001]
Negative Affect	0.98	8.21 (4, 19,868)	< .001*	.001	[.000, .001]
Loneliness	0.94	4.87 (4, 19,860)	.001	.000	[.000, .001]
Organizational Trust	0.96	141.60 (4, 19,864)	< .001*	.015	[.013, .018]
Work Engagement	0.96	210.07 (4, 19,860)	< .001*	.016	[.016, .022]
Life Meaning	0.98	351.78 (4, 19,864)	< .001*	.028	[.024, .032]

Note. GAT = Global Assessment Tool; Model *n*s = 4,966–4,968; $\varepsilon_{G\text{-}G}$ = Greenhouse-Geisser epsilon estimate of sphericity; $^*p < .05$ after applying family-wise Bonferroni correction ($\alpha/130 = .0004$); Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); CI = confidence interval.

Table E2

Table of Repeated Measures ANOVA (RM-ANOVA) Tests Across Five Time Occasions for Each GAT 2.0 Measure

Measure	$\varepsilon_{G\text{-}G}$	F (df)	p	η^2	90% CI η^2
Adaptability	0.96	2.66 (4, 2,628)	.031	.002	[.000, .003]
Active Coping	0.98	4.36 (4, 2,628)	.002	.002	[.000, .005]
Passive Coping	0.98	2.77 (4, 2,628)	.026	.002	[.000, .005]
Character	0.94	7.50 (4, 2,624)	< .001*	.004	[.000, .007]
Catastrophizing	0.94	4.63 (4, 2,628)	.001	.003	[.000, .006]
Depression	0.97	7.68 (4, 2,624)	< .001*	.005	[.001, .009]
Optimism	0.93	0.88 (4, 564)	.476	.002	[.000, .005]
Positive Affect	0.90	1.40 (4, 564)	.233	.003	[.000, .007]
Negative Affect	0.93	1.34 (4, 564)	.252	.004	[.000, .009]
Loneliness	0.93	1.09 (4, 2,628)	.359	.001	[.000, .001]
Organizational Trust	0.94	29.70 (4, 2,628)	< .001*	.021	[.012, .030]
Work Engagement	0.89	26.96 (4, 2,628)	< .001*	.016	[.008, .024]
Life Meaning	0.94	10.00 (4, 2,628)	< .001*	.005	[.001, .009]

Note. GAT = Global Assessment Tool; Model *n*s = 142–658; $\varepsilon_{G\text{-}G}$ = Greenhouse-Geisser epsilon estimate of sphericity; $^*p < .05$ after applying family-wise Bonferroni correction ($\alpha/130 = .0004$); Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); CI = confidence interval.

Appendix F: Repeated Measures Structural Equation Model (RM-SEM) Results

Table F1

Table of Repeated Measures SEM (RM-SEM) Tests Across Five Time Occasions for Each GAT 1.0 Measure

		Model Fit Statistics						Model Comparisons					Measures of Effect Size		
Measure	Model	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI_{RMSEA}	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Full	1,675.09 (13)	.964	.979	.032	[.031, .034]	.061	—	—	—	—	—	—	—	—
	M1b: Null	1,696.21 (17)	.964	.972	.037	[.035, .038]	.060	M1a	21.12 (4)*	.000	−.007	.005	0.007	.000	[.000, .000]
Active Coping	M2a: Full	3,417.43 (13)	.931	.947	.052	[.051, .054]	.096	—	—	—	—	—	—	—	—
	M2b: Null	3,575.89 (17)	.928	.957	.047	[.046, .048]	.097	M2a	158.46 (4)*	−.003	.010	−.005	0.020	.001	[.001, .001]
Passive Coping	M3a: Full	903.76 (13)	.975	.981	.027	[.025, .028]	.049	—	—	—	—	—	—	—	—
	M3b: Null	2,264.93 (17)	.937	.963	.037	[.036, .039]	.062	M3a	1,361.17 (4)*	−.038	.018	−.010	0.060	.008	[.007, .008]
Character	M4a: Full	6,554.48 (13)	.871	.900	.073	[.071, .074]	.135	—	—	—	—	—	—	—	—
	M4b: Null	6,591.70 (17)	.870	.923	.064	[.062, .065]	.135	M4a	37.22 (4)*	−.001	.023	−.009	0.010	.000	[.000, .000]
Catastrophizing	M5a: Full	2,404.38 (13)	.943	.956	.044	[.042, .045]	.084	—	—	—	—	—	—	—	—
	M5b: Null	2,994.30 (17)	.928	.958	.043	[.042, .044]	.086	M5a	589.93 (4)*	−.015	.002	−.001	0.039	.003	[.003, .004]
Depression	M6a: Full	3,532.08 (13)	.901	.924	.053	[.052, ..055]	.095	—	—	—	—	—	—	—	—
	M6b: Null	3,980.32 (17)	.888	.934	.049	[.048, .051]	.098	M6a	448.24 (4)*	−.013	.010	−.004	0.034	.003	[.002, .003]
Optimism	M7a: Full	2,208.40 (13)	.963	.972	.042	[.041, .044]	.067	—	—	—	—	—	—	—	—
	M7b: Null	2,449.93 (17)	.959	.976	.039	[.037, .040]	.067	M7a	241.52 (4)*	−.004	.004	−.003	0.025	.001	[.001, .002]
Positive Affect	M8a: Full	3,470.49 (13)	.937	.951	.053	[.051, .054]	.094	—	—	—	—	—	—	—	—
	M8b: Null	3,558.09 (17)	.935	.962	.047	[.045, .048]	.095	M8a	87.60 (4)*	−.002	.011	−.006	0.015	.000	[.000, .001]
Negative Affect	M9a: Full	2,524.46 (13)	.941	.955	.045	[.044, .047]	.077	—	—	—	—	—	—	—	—
	M9b: Null	2,605.04 (17)	.940	.964	.040	[.039, .041]	.077	M9a	80.58 (4)*	−.001	.009	−.005	0.015	.000	[.000, .001]
Loneliness	M10a: Full	2,215.80 (13)	.961	.970	.042	[.041, .044]	.072	—	—	—	—	—	—	—	—
	M10b: Null	2,225.99 (17)	.961	.977	.037	[.036, .038]	.072	M10a	10.19 (4)	.000	.007	−.005	0.005	.000	[.000, .000]
Organizational Trust	M11a: Full	6,376.37 (13)	.747	.806	.072	[.070, .073]	.126	—	—	—	—	—	—	—	—
	M11b: Null	12,472.71 (17)	.506	.709	.088	[.086, .089]	.147	M11a	6,096.34 (4)*	−.241	−.097	.016	0.126	.033	[.032, .035]
Work Engagement	M12a: Full	5,913.11 (13)	.864	.895	.069	[.068, .071]	.120	—	—	—	—	—	—	—	—
	M12b: Null	15,234.48 (17)	.648	.793	.097	[.096, .098]	.143	M12a	9,321.37 (4)*	−.216	−.102	.028	0.156	.050	[.048, .052]
Life Meaning	M13a: Full	3,195.99 (13)	.946	.959	.051	[.049, .052]	.074	—	—	—	—	—	—	—	—
	M13b: Null	9,578.14 (17)	.838	.905	.077	[.076, .078]	.133	M13a	6,382.15 (4)*	−.108	−.054	.026	0.129	.035	[.033, .036]

Note. GAT = Global Assessment Tool; Model *n*s = 95,277; All $\Delta\chi^2$ *p*s < .05, **p* < .05 after applying family-wise Bonferroni correction (α/130 = .0004); *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Table F2

Table of Repeated Measures SEM (RM-SEM) Tests Across Five Time Occasions for Each GAT 2.0 Measure

		Model Fit Statistics						Model Comparisons					Measures of Effect Size		
Measure	Model	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI_{RMSEA}	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Full	588.46 (13)	.974	.980	.028	[.026, .030]	.064	—	—	—	—	—	—	—	—
	M1b: Null	1,289.77 (17)	.943	.966	.036	[.034, .038]	.076	M1a	701.31 (4)*	−.031	−.014	.008	0.055	.008	[.007, .009]
Active Coping	M2a: Full	1,214.32 (13)	.950	.961	.040	[.038, .042]	.089	—	—	—	—	—	—	—	—
	M2b: Null	1,818.68 (17)	.925	.956	.043	[.041, .045]	.099	M2a	604.36 (4)*	−.025	−.005	.003	0.051	.007	[.006, .008]
Passive Coping	M3a: Full	191.04 (13)	.987	.990	.015	[.014, .017]	.041	—	—	—	—	—	—	—	—
	M3b: Null	319.81 (17)	.979	.987	.018	[.016, .019]	.050	M3a	128.77 (4)*	−.008	−.003	.003	0.024	.002	[.001, .002]
Character	M4a: Full	3,293.42 (13)	.875	.904	.066	[.064, .068]	.146	—	—	—	—	—	—	—	—
	M4b: Null	3,833.23 (17)	.855	.915	.062	[.061, .064]	.154	M4a	539.82 (4)*	−.020	.011	−.004	0.048	.007	[.006, .007]
Catastrophizing	M5a: Full	952.31 (13)	.943	.956	.035	[.033, .037]	.076	—	—	—	—	—	—	—	—
	M5b: Null	1,877.44 (17)	.887	.933	.044	[.042, .045]	.086	M5a	925.13 (4)*	−.056	−.023	.009	0.063	.011	[.010, .012]
Depression	M6a: Full	2,681.20 (13)	.837	.875	.060	[.058, .062]	.136	—	—	—	—	—	—	—	—
	M6b: Null	3,860.64 (17)	.766	.862	.063	[.061, .064]	.147	M6a	1,179.44 (4)*	−.071	−.013	.003	0.071	.014	[.013, .015]
Optimism	M7a: Full	544.51 (13)	.968	.975	.027	[.025, .029]	.104	—	—	—	—	—	—	—	—
	M7b: Null	856.37 (17)	.949	.970	.029	[.028, .031]	.108	M7a	311.85 (4)*	−.019	−.005	.002	0.037	.007	[.005, .008]
Positive Affect	M8a: Full	945.90 (13)	.937	.952	.035	[.033, .037]	.116	—	—	—	—	—	—	—	—
	M8b: Null	1,532.37 (17)	.898	.940	.039	[.038, .041]	.128	M8a	586.46 (4)*	−.039	−.012	.004	0.050	.012	[.011, .014]
Negative Affect	M9a: Full	442.07 (13)	.961	.970	.024	[.022, .026]	.087	—	—	—	—	—	—	—	—
	M9b: Null	538.00 (17)	.953	.972	.023	[.021, .025]	.093	M9a	95.93 (4)*	−.008	.002	−.001	0.020	.002	[.001, .003]
Loneliness	M10a: Full	851.38 (13)	.969	.976	.033	[.032, .035]	.075	—	—	—	—	—	—	—	—
	M10b: Null	1,144.52 (17)	.959	.976	.034	[.032, .036]	.080	M10a	293.14 (4)*	−.010	.000	.001	0.036	.004	[.003, .004]
Organizational Trust	M11a: Full	3,175.66 (13)	.742	.801	.065	[.063, .067]	.129	—	—	—	—	—	—	—	—
	M11b: Null	7,094.73 (17)	.422	.660	.085	[.083, .087]	.162	M11a	3,919.06 (4)*	−.320	−.141	.020	0.130	.045	[.043, .048]
Work Engagement	M12a: Full	2,655.35 (13)	.873	.902	.059	[.057, .061]	.133	—	—	—	—	—	—	—	—
	M12b: Null	8,917.55 (17)	.573	.749	.095	[.094, .097]	.171	M12a	6,262.19 (4)*	−.300	−.153	.036	0.165	.071	[.068, .074]
Life Meaning	M13a: Full	2,836.63 (13)	.897	.921	.061	[.059, .063]	.138	—	—	—	—	—	—	—	—
	M13b: Null	5,140.55 (17)	.813	.890	.072	[.071, .074]	.154	M13a	2,303.92 (4)*	−.084	−.031	.011	0.100	.027	[.025, .029]

Note. GAT = Global Assessment Tool; Model *n*s = 57,495–57,771; All $\Delta\chi^2$ *p*s < .001, **p* < .05 after applying family-wise Bonferroni correction (α/130 = .0004); *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Appendix G: Measurement Invariance (MI) Results

Table G1

Table of Measurement Invariance (MI) Tests Across Five Time Occasions for Each GAT 1.0 Measure

		Model Fit Statistics						Model Comparison					Measures of Effect Size		
Measure	Model	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI_{RMSEA}	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Configural	7,633.69 (68)	.971	.955	.034	[.034, .035]	.042	—	—	—	—	—	—	—	—
	M1b: Weak (λ = λ)	8,104.96 (76)	.969	.958	.033	[.033, .034]	.047	M1a	471.27 (8)*	−.002	.003	−.001	0.025	.001	[.001, .001]
	M1c: Strong ($\tau = \tau$)	8,240.99 (84)	.969	.961	.032	[.031, .033]	.048	M1b	136.03 (8)*	.000	.003	−.001	0.013	.000	[.000, .000]
Active Coping	M2a: Configural	36,940.13 (245)	.928	.911	.040	[.039, .040]	.037	—	—	—	—	—	—	—	—
	M2b: Weak (λ = λ)	37,091.44 (261)	.927	.917	.038	[.038, .039]	.038	M2a	151.31 (16)*	−.001	.006	−.002	0.010	.000	[.000, .000]
	M2c: Strong ($\tau = \tau$)	38,179.79 (277)	.925	.919	.038	[.038, .038]	.038	M2b	1,088.35 (16)*	−.002	.002	.000	0.027	.001	[.001, .001]
Passive Coping	M3a: Configural	9,368.01 (68)	.967	.949	.038	[.037, .039]	.055	—	—	—	—	—	—	—	—
	M3b: Weak (λ = λ)	9,639.41 (76)	.966	.953	.036	[.036, .037]	.058	M3a	271.40 (8)*	−.001	.004	−.002	0.019	.000	[.000, .000]
	M3c: Strong ($\tau = \tau$)	9,681.77 (84)	.966	.958	.034	[.034, .035]	.057	M3b	42.36 (8)*	.000	.005	−.002	0.007	.000	[.000, .000]
Character	M4a: Configural	887,156.65 (6,914)	.846	.841	.037	[.036, .037]	.032	—	—	—	—	—	—	—	—
	M4b: Weak (λ = λ)	888,891.32 (7,006)	.846	.843	.036	[.036, .036]	.032	M4a	1,734.68 (92)*	.000	.002	−.001	0.014	.000	[.000, .000]
	M4c: Strong ($\tau = \tau$)	924,647.94 (7,098)	.840	.839	.037	[.037, .037]	.034	M4b	35,756.62 (92)*	−.006	−.004	.001	0.064	.006	[.005, .006]
Catastrophizing	M5a: Configural	94,241.55 (522)	.902	.888	.043	[.043, .044]	.041	—	—	—	—	—	—	—	—
	M5b: Weak (λ = λ)	94,952.60 (546)	.901	.892	.043	[.042, .043]	.041	M5a	711.05 (24)*	−.001	.004	.000	0.018	.000	[.000, .000]
	M5c: Strong ($\tau = \tau$)	101,186.94 (570)	.895	.890	.043	[.043, .043]	.042	M5b	6,234.34 (24)*	−.006	−.002	.000	0.052	.003	[.003, .004]
Depression	M6a: Configural	182,019.56 (1,125)	.902	.894	.041	[.041, .041]	.036	—	—	—	—	—	—	—	—
	M6b: Weak (λ = λ)	183,409.77 (1,161)	.902	.896	.041	[.040, .041]	.038	M6a	1,390.20 (36)*	.000	.002	.000	0.020	.000	[.000, .001]
	M6c: Strong ($\tau = \tau$)	189,019.30 (1,197)	.899	.896	.041	[.040, .041]	.038	M6b	5,609.53 (36)*	−.003	.000	.000	0.040	.002	[.002, .002]
Optimism	M7a: Configural	50,277.61 (144)	.858	.813	.060	[.060, .061]	.053	—	—	—	—	—	—	—	—
	M7b: Weak (λ = λ)	50,625.00 (156)	.858	.826	.058	[.058, .059]	.056	M7a	347.40 (12)*	.000	.013	−.002	0.017	.000	[.000, .000]
	M7c: Strong ($\tau = \tau$)	51,120.78 (168)	.856	.837	.056	[.056, .057]	.056	M7b	495.77 (12)*	−.002	.011	−.002	0.021	.000	[.000, .001]
Positive Affect	M8a: Configural	120,909.35 (1,125)	.943	.938	.033	[.033, .034]	.025	—	—	—	—	—	—	—	—
	M8b: Weak (λ = λ)	121,751.52 (1,161)	.942	.939	.033	[.033, .033]	.026	M8a	842.17 (36)*	−.001	.001	.000	0.016	.000	[.000, .000]
	M8c: Strong ($\tau = \tau$)	137,254.35 (1,197)	.935	.934	.035	[.034, .035]	.028	M8b	15,502.82 (36)*	−.007	−.005	.002	0.067	.006	[.006, .006]
Negative Affect	M9a: Configural	241,523.24 (1,376)	.850	.838	.043	[.043, .043]	.046	—	—	—	—	—	—	—	—
	M9b: Weak (λ = λ)	242,876.00 (1,416)	.849	.842	.042	[.042, .042]	.048	M9a	1,352.76 (40)*	−.001	.004	−.001	0.019	.000	[.000, .000]
	M9c: Strong ($\tau = \tau$)	259,231.66 (1,456)	.839	.836	.043	[.043, .043]	.049	M9b	16,355.66 (40)*	−.010	−.006	.001	0.066	.006	[.005, .006]
Loneliness	M10a: Configural	16,757.39 (68)	.957	.934	.051	[.050, .051]	.077	—	—	—	—	—	—	—	—
	M10b: Weak (λ = λ)	16,865.37 (76)	.957	.941	.048	[.048, .049]	.077	M10a	107.98 (8)*	.000	.007	−.003	0.012	.000	[.000, .000]
	M10c: Strong ($\tau = \tau$)	18,442.48 (84)	.953	.941	.048	[.047, .048]	.078	M10b	1,577.11 (8)*	−.004	.000	.000	0.045	.002	[.002, .002]
Organizational Trust	M11a: Configural	62,206.72 (245)	.909	.888	.052	[.051, .052]	.049	—	—	—	—	—	—	—	—
	M11b: Weak (λ = λ)	62,407.86 (261)	.908	.895	.050	[.050, .050]	.051	M11a	201.14 (16)*	−.001	.007	−.002	0.011	.000	[.000, .000]
	M11c: Strong ($\tau = \tau$)	73,360.13 (277)	.892	.883	.053	[.052, .053]	.055	M11b	10,952.26 (16)*	−.016	−.012	.003	0.085	.009	[.008, .009]
Work Engagement	M12a: Configural	22,096.72 (144)	.955	.941	.040	[.040, .040]	.034	—	—	—	—	—	—	—	—
	M12b: Weak (λ = λ)	22,199.05 (156)	.955	.945	.039	[.038, .039]	.034	M12a	102.33 (12)*	.000	.004	−.001	0.009	.000	[.000, .000]
	M12c: Strong ($\tau = \tau$)	22,520.88 (168)	.954	.948	.037	[.037, .038]	.034	M12b	321.82 (12)*	−.001	.003	−.002	0.017	.000	[.000, .000]
Life Meaning	M13a: Configural	53,679.79 (245)	.925	.908	.048	[.048, .048]	.058	—	—	—	—	—	—	—	—
	M13b: Weak (λ = λ)	56,534.79 (261)	.921	.909	.048	[.047, .048]	.070	M13a	2,855.00 (16)*	−.004	.001	.000	0.043	.002	[.002, .002]
	M13c: Strong ($\tau = \tau$)	100,545.82 (277)	.859	.847	.062	[.061, .062]	.104	M13b	44,011.02 (16)*	−.062	−.062	.014	0.170	.034	[.033, .034]

Note. GAT = Global Assessment Tool; Model *n*s = 95,277; All $\Delta\chi^2$ *p*s < .001, *p < .05 after applying family-wise Bonferroni correction (α/130 = .0004); *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Table G2

Table of Measurement Invariance (MI) Tests Across Five Time Occasions for Each GAT 2.0 Measure

		Model Fit Statistics						Model Comparison					Measures of Effect Size		
Measure	Model	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI_{RMSEA}	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Configural	3,183.92 (68)	.975	.961	.028	[.027, .029]	.050	—	—	—	—	—	—	—	—
	M1b: Weak (λ = λ)	3,376.53 (76)	.973	.963	.027	[.027, .028]	.052	M1a	192.60 (8)*	−.002	.002	−.001	0.020	.001	[.000, .001]
	M1c: Strong ($\tau = \tau$)	4,102.55 (84)	.967	.959	.028	[.028, .030]	.053	M1b	726.03 (8)*	−.006	−.004	.001	0.040	.002	[.002, .002]
Active Coping	M2a: Configural	13,379.57 (245)	.948	.937	.030	[.030, .031]	.041	—	—	—	—	—	—	—	—
	M2b: Weak (λ = λ)	13,433.30 (261)	.948	.941	.030	[.029, .030]	.042	M2a	53.73 (16)*	.000	.004	.000	0.008	.000	[.000, .000]
	M2c: Strong ($\tau = \tau$)	13,896.66 (277)	.947	.942	.029	[.029, .030]	.042	M2b	463.36 (16)*	−.001	.001	−.001	0.022	.001	[.001, .001]
Passive Coping	M3a: Configural	2,686.71 (68)	.980	.969	.026	[.025, .027]	.059	—	—	—	—	—	—	—	—
	M3b: Weak (λ = λ)	2,744.20 (76)	.980	.972	.025	[.024, .025]	.060	M3a	57.49 (8)*	.000	.003	−.001	0.011	.000	[.000, .000]
	M3c: Strong ($\tau = \tau$)	2,847.07 (84)	.979	.974	.024	[.023, .025]	.060	M3b	102.87 (8)*	−.001	.002	−.001	0.015	.000	[.000, .000]
Character	M4a: Configural	275,564.19 (3,833)	.865	.859	.035	[.035, .035]	.036	—	—	—	—	—	—	—	—
	M4b: Weak (λ = λ)	276,559.34 (3,901)	.865	.861	.035	[.035, .035]	.037	M4a	995.15 (68)*	.000	.002	.000	0.016	.000	[.000, .000]
	M4c: Strong ($\tau = \tau$)	291,262.83 (3,969)	.858	.856	.035	[.035, .036]	.038	M4b	14,703.49 (68)*	−.007	−.005	.000	0.061	.006	[.006, .006]
Catastrophizing	M5a: Configural	3,664.19 (68)	.974	.960	.030	[.029, .031]	.035	—	—	—	—	—	—	—	—
	M5b: Weak (λ = λ)	3,777.57 (76)	.974	.964	.029	[.028, .030]	.038	M5a	113.38 (8)*	.000	.004	−.001	0.016	.000	[.000, .000]
	M5c: Strong ($\tau = \tau$)	4,231.48 (84)	.970	.963	.029	[.028, .030]	.040	M5b	453.91 (8)*	−.004	−.001	.000	0.031	.001	[.001, .001]
Depression	M6a: Configural	86,116.38 (1,125)	.909	.900	.036	[.036, .036]	.040	—	—	—	—	—	—	—	—
	M6b: Weak (λ = λ)	86,778.72 (1,161)	.908	.903	.036	[.036, .036]	.043	M6a	662.34 (36)*	−.001	.003	.000	0.018	.000	[.000, .001]
	M6c: Strong ($\tau = \tau$)	89,074.83 (1,197)	.906	.903	.036	[.035, .036]	.043	M6b	2,296.11 (36)*	−.002	.000	.000	0.033	.002	[.002, .002]
Optimism	M7a: Configural	20,614.35 (144)	.832	.779	.050	[.049, .050]	.081	—	—	—	—	—	—	—	—
	M7b: Weak (λ = λ)	20,764.43 (156)	.831	.794	.048	[.047, .048]	.086	M7a	150.08 (12)*	−.001	.015	−.002	0.015	.000	[.000, .001]
	M7c: Strong ($\tau = \tau$)	20,852.33 (168)	.830	.808	.046	[.046, .047]	.085	M7b	87.90 (12)*	−.001	.014	−.002	0.011	.000	[.000, .000]
Positive Affect	M8a: Configural	36,393.51 (899)	.948	.943	.026	[.026, .026]	.070	—	—	—	—	—	—	—	—
	M8b: Weak (λ = λ)	37,069.55 (931)	.947	.944	.026	[.026, .026]	.073	M8a	676.04 (32)*	−.001	.001	.000	0.019	.001	[.001, .001]
	M8c: Strong ($\tau = \tau$)	44,727.46 (963)	.936	.934	.028	[.028, .028]	.074	M8b	7,657.92 (32)*	−.011	−.010	.002	0.065	.009	[.008, .009]
Negative Affect	M9a: Configural	46,794.61 (899)	.896	.886	.030	[.030, .030]	.069	—	—	—	—	—	—	—	—
	M9b: Weak (λ = λ)	47,215.24 (931)	.895	.889	.029	[.029, .030]	.069	M9a	420.63 (32)*	−.001	.003	−.001	0.015	.000	[.000, .001]
	M9c: Strong ($\tau = \tau$)	50,773.96 (963)	.887	.884	.030	[.030, .030]	.070	M9b	3,558.72 (32)*	−.008	−.005	.001	0.044	.004	[.004, .004]
Loneliness	M10a: Configural	6,760.30 (68)	.964	.944	.041	[.040, .042]	.086	—	—	—	—	—	—	—	—
	M10b: Weak (λ = λ)	6,847.85 (76)	.963	.949	.039	[.038, .040]	.088	M10a	87.55 (8)*	−.001	.005	−.002	0.014	.000	[.000, .000]
	M10c: Strong ($\tau = \tau$)	7,229.00 (84)	.961	.952	.038	[.038, .039]	.090	M10b	381.15 (8)*	−.002	.003	−.001	0.029	.001	[.001, .001]
Organizational Trust	M11a: Configural	12,832.25 (144)	.957	.944	.039	[.038, .040]	.054	—	—	—	—	—	—	—	—
	M11b: Weak (λ = λ)	12,944.22 (156)	.957	.947	.038	[.037, .038]	.055	M11a	111.97 (12)*	.000	.003	−.001	0.013	.000	[.000, .000]
	M11c: Strong ($\tau = \tau$)	15,238.07 (168)	.949	.943	.039	[.039, .040]	.058	M11b	2,293.85 (12)*	−.008	−.004	.001	0.058	.005	[.004, .005]
Work Engagement	M12a: Configural	7,096.82 (144)	.973	.964	.029	[.028, .029]	.039	—	—	—	—	—	—	—	—
	M12b: Weak (λ = λ)	7,197.66 (156)	.973	.967	.028	[.027, .029]	.041	M12a	100.84 (12)*	.000	.003	−.001	0.012	.000	[.000, .000]
	M12c: Strong ($\tau = \tau$)	7,372.48 (168)	.972	.968	.027	[.027, .028]	.042	M12b	174.82 (12)*	−.001	.001	−.001	0.016	.000	[.000, .000]
Life Meaning	M13a: Configural	20,339.17 (245)	.952	.941	.038	[.037, .038]	.053	—	—	—	—	—	—	—	—
	M13b: Weak (λ = λ)	20,404.23 (261)	.952	.945	.037	[.036, .037]	.055	M13a	65.07 (16)*	.000	.004	−.001	0.008	.000	[.000, .000]
	M13c: Strong ($\tau = \tau$)	25,760.62 (277)	.939	.934	.040	[.039, .040]	.059	M13b	5,356.39 (16)*	−.013	−.011	.003	0.076	.008	[.008, .009]

Note. GAT = Global Assessment Tool; Model *n*s = 57,495–57,771; All $\Delta\chi^2$ *p*s < .001, **p* < .05 after applying family-wise Bonferroni correction (α/130 = .0004); *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker−Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Appendix H: Repeated Measures Confirmatory Factor Analysis (RM-CFA) Results

Table H1

Table of Repeated Measures CFA (RM-CFA) Tests Across Five Time Occasions for Each GAT 1.0 Measure

		Model Fit Statistics						Model Comparison					Measures of Effect Size		
Measure	Model	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI_{RMSEA}	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Full	10,220.54 (97)	.961	.958	.033	[.033, .034]	.072	—	—	—	—	—	—	—	—
	M1b: Null	10,249.20 (101)	.961	.960	.032	[.032, .033]	.071	M1a	28.67 (4)*	.000	.002	−.001	0.009	.000	[.000, .000]
Active Coping	M2a: Full	42,420.77 (290)	.917	.914	.039	[.039, .039]	.078	—	—	—	—	—	—	—	—
	M2b: Null	42,590.62 (294)	.917	.914	.039	[.039, .039]	.079	M2a	169.85 (4)*	.000	.000	.000	0.021	.000	[.000, .000]
Passive Coping	M3a: Full	10,561.91 (97)	.963	.960	.034	[.033,.034]	.065	—	—	—	—	—	—	—	—
	M3b: Null	12,062.56 (101)	.958	.956	.035	[.035, .036]	.065	M3a	1,500.65 (4)*	−.005	−.004	.001	0.063	.002	[.002, .002]
Character	M4a: Full	931,371.90 (7,111)	.839	.838	.037	[.037, .037]	.082	—	—	—	—	—	—	—	—
	M4b: Null	931,450.58 (7,115)	.839	.838	.037	[.037, .037]	.082	M4a	78.68 (4)*	.000	.000	.000	0.014	.000	[.000, .000]
Catastrophizing	M5a: Full	103,974.29 (583)	.892	.889	.043	[.043, .043]	.064	—	—	—	—	—	—	—	—
	M5b: Null	104,273.15 (587)	.891	.890	.043	[.043, .043]	.066	M5a	298.86 (4)*	−.001	.001	.000	0.028	.000	[.000, .000]
Depression	M6a: Full	192,589.06 (1,210)	.897	.895	.041	[.041, .041]	.068	—	—	—	—	—	—	—	—
	M6b: Null	193,010.88 (1,214)	.896	.895	.041	[.041, .041]	.068	M6a	421.82 (4)*	−.001	.000	.000	0.033	.000	[.000, .000]
Optimism	M7a: Full	53,842.22 (181)	.849	.841	.056	[.055, .056]	.077	—	—	—	—	—	—	—	—
	M7b: Null	5,4052.50 (185)	.848	.844	.055	[.055, .056]	.077	M7a	210.27 (4)*	−.001	.003	−.001	0.023	.000	[.000, .000]
Positive Affect	M8a: Full	14,0961.80 (1,210)	.933	.932	.035	[.035, .035]	.071	—	—	—	—	—	—	—	—
	M8b: Null	141,129.07 (1,214)	.933	.933	.035	[.035, .035]	.071	M8a	167.27 (4)*	.000	.001	.000	0.021	.000	[.000, .000]
Negative Affect	M9a: Full	261,852.60 (1,469)	.837	.835	.043	[.043, .043]	.066	—	—	—	—	—	—	—	—
	M9b: Null	261,867.03 (1,473)	.837	.836	.043	[.043, .043]	.066	M9a	14.43 (4)	.000	.001	.000	0.006	.000	[.000, .000]
Loneliness	M10a: Full	20,548.89 (97)	.948	.943	.047	[.047, .048]	.099	—	—	—	—	—	—	—	—
	M10b: Null	20,597.83 (101)	.947	.945	.046	[.046, .047]	.098	M10a	48.94 (4)*	−.001	.002	−.001	0.011	.000	[.000, .000]
Organizational Trust	M11a: Full	79,862.38 (290)	.883	.879	.054	[.053, .054]	.109	—	—	—	—	—	—	—	—
	M11b: Null	87,630.69 (294)	.871	.869	.056	[.056, .056]	.132	M11a	7,768.31 (4)*	−.012	−.010	.002	0.143	.006	[.006, .006]
Work Engagement	M12a: Full	29,064.22 (181)	.941	.938	.041	[.041, .041]	.098	—	—	—	—	—	—	—	—
	M12b: Null	39,313.03 (185)	.920	.918	.047	[.047, .048]	.121	M12a	10,248.81 (4)*	−.021	−.020	.006	0.164	.010	[.010, .011]
Life Meaning	M13a: Full	103,072.88 (290)	.856	.851	.061	[.061, .061]	.113	—	—	—	—	—	—	—	—
	M13b: Null	107,051.68 (294)	.850	.847	.062	[.061, .062]	.125	M13a	3,978.80 (4)*	−.006	−.004	.001	0.102	.003	[.003, .003]

Note. GAT = Global Assessment Tool; Model *n*s = 95,277; All $\Delta\chi^2$ *p*s < .01, *p < .05 after applying family-wise Bonferroni correction (α/130 = .0004); *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Table H2

Table of Repeated Measures CFA (RM-CFA) Tests Across Five Time Occasions for Each GAT 2.0 Measure

		Model Fit Statistics						Model Comparison					Measures of Effect Size		
Measure	Model	χ^2 (*df*)	CFI	TLI	RMSEA	90% CI_{RMSEA}	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Full	4,799.54 (97)	.962	.959	.029	[.028, .030]	.083	—	—	—	—	—	—	—	—
	M1b: Null	5,136.54 (101)	.959	.957	.029	[.029, .030]	.091	M1a	337.01 (4)*	−.003	−.002	.000	0.038	.001	[.001, .001]
Active Coping	M2a: Full	15,464.08 (290)	.940	.938	.030	[.030, .030]	.079	—	—	—	—	—	—	—	—
	M2b: Null	16,141.28 (294)	.938	.937	.031	[.030, .031]	.089	M2a	677.20 (4)*	−.002	−.001	.001	0.054	.001	[.001, .001]
Passive Coping	M3a: Full	3,054.58 (97)	.977	.976	.023	[.022, .024]	.067	—	—	—	—	—	—	—	—
	M3b: Null	3,241.86 (101)	.976	.975	.023	[.023, .024]	.067	M3a	187.28 (4)*	−.001	−.001	.000	0.028	.001	[.000, .001]
Character	M4a: Full	294,668.50 (3,982)	.856	.855	.036	[.035, .036]	.091	—	—	—	—	—	—	—	—
	M4b: Null	295,354.69 (3,986)	.856	.855	.036	[.035, .036]	.097	M4a	686.20 (4)*	.000	.000	.000	0.054	.000	[.000, .000]
Catastrophizing	M5a: Full	5,525.46 (97)	.961	.958	.031	[.030, .032]	.070	—	—	—	—	—	—	—	—
	M5b: Null	6,613.40 (101)	.954	.952	.033	[.033, .034]	.067	M5a	1,087.94 (4)*	−.007	−.006	.002	0.069	.003	[.003, .003]
Depression	M6a: Full	91,787.19 (1,210)	.903	.901	.036	[.036, .036]	.086	—	—	—	—	—	—	—	—
	M6b: Null	92,892.82 (1,214)	.901	.901	.036	[.036, .036]	.086	M6a	1,105.64 (4)*	−.002	.000	.000	0.069	.001	[.001, .001]
Optimism	M7a: Full	21,526.60 (181)	.825	.816	.045	[.045, .046]	.160	—	—	—	—	—	—	—	—
	M7b: Null	21,831.22 (185)	.823	.818	.045	[.045, .046]	.112	M7a	304.61 (4)*	−.002	.002	.000	0.036	.001	[.001, .001]
Positive Affect	M8a: Full	45,723.62 (976)	.935	.934	.028	[.028, .028]	.097	—	—	—	—	—	—	—	—
	M8b: Null	46,345.13 (980)	.934	.933	.028	[.028, .029]	.106	M8a	621.55 (4)*	−.001	−.001	.000	0.052	.001	[.001, .001]
Negative Affect	M9a: Full	51,263.95 (976)	.886	.885	.030	[.030, .030]	.068	—	—	—	—	—	—	—	—
	M9b: Null	51,398.38 (980)	.886	.885	.030	[.030, .030]	.066	M9a	134.42 (4)*	.000	.000	.000	0.024	.000	[.000, .000]
Loneliness	M10a: Full	8,106.44 (97)	.957	.953	.038	[.037, .039]	.110	—	—	—	—	—	—	—	—
	M10b: Null	8,604.91 (101)	.954	.952	.038	[.037, .039]	.105	M10a	498.47 (4)*	−.003	−.001	.000	0.046	.001	[.001, .002]
Organizational Trust	M11a: Full	18,451.58 (181)	.938	.935	.042	[.041, .042]	.131	—	—	—	—	—	—	—	—
	M11b: Null	22,823.77 (185)	.924	.922	.046	[.046, .047]	.167	M11a	4,372.19 (4)*	−.014	−.013	.004	0.138	.009	[.008, .009]
Work Engagement	M12a: Full	10,340.96 (181)	.960	.959	.031	[.031, .032]	.114	—	—	—	—	—	—	—	—
	M12b: Null	17,202.07 (185)	.934	.932	.040	[.039, .040]	.148	M12a	6,861.11 (4)*	−.026	−.027	.009	0.172	.013	[.013, .014]
Life Meaning	M13a: Full	28,767.23 (290)	.932	.930	.041	[.041, .042]	.143	—	—	—	—	—	—	—	—
	M13b: Null	31,045.59 (294)	.927	.925	.043	[.042, .043]	.163	M13a	2,278.35 (4)*	−.005	−.005	.002	0.099	.004	[.003, .004]

Note. GAT = Global Assessment Tool; Model *n*s = 57,495–57,771; All $\Delta\chi^2$ *p*s < .001, *p < .05 after applying family-wise Bonferroni correction (α/130 = .0004); *df* = degrees of freedom; CFI = comparative fit index; TLI = Tucker−Lewis index; RMSEA = root mean square error of approximation; SRMR = standardized root mean square residual; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Appendix I: Repeated Measures Multi-Level Model (RM-MLM) Results

In addition, as an extension to the models described above which treated time occasions as categorical, we also examined time as a continuous metric. Treating time as an individually-varying metric allowed us to examine our research questions in a way that accounts for unequal spacing of assessment dates for each respondent in our samples. Using a repeated measures multi-level modeling framework (RM-MLM), we were able to treat time as a continuous metric for each respondent. The RM-MLM framework also allowed us to obtained estimates of the random intercepts and random slopes for each of the GAT measures. Estimates of random intercepts indicate the degree to which there is individual variability in the starting point for each respondent (i.e., value at time zero). Estimates of random slopes indicate the degree to which there is individual variability in the slope across time (i.e., the pattern in which values change over time). Thus, whereas RM-ANOVA shows aggregate (i.e., averaged across individuals) change over time across individuals, comparisons of random slopes allow us to check to see if there is a degree of within-individual change. Stated differently, random slopes show how closely the patterns of individuals across time match the aggregate fixed effect of time.

To examine these estimates, we conducted two nested models using the nlme multi-level modeling (linear mixed effects models) package in R (version 3.1.14) for the GAT 1.0 data in Study 1 and the GAT 2.0 data in Study 2. For these models, the scale-level composites (aggregated mean across individual items) of the 13 measures were used in separate tests. The five time occasions were modeled continuously instead of categorically for each respondent. Using the completion dates for each time occasion, we calculated the amount of time that had passed (in years) since a respondent first took the respective version of the GAT [Study 1 *M* (*SD*): T1 = 0.00 (0.00), T2 = 0.90 (0.33), T3 = 1.68 (0.42), T4 = 2.48 (0.46), T5 = 3.29 (0.48); Study 2 *M* (*SD*): T1 = 0.00 (0.00), T2 = 0.66 (0.34), T3 = 1.30 (0.48), T4 = 1.97 (0.57), T5 = 2.71 (0.59)]. RM-MLMs were estimated with maximum likelihood (ML) and missing data were treated with listwise deletion. Covariance matrices were structured using compound symmetry such that all variances and covariances were set to be equal across time.[7] In the first (full) model, the random intercept and random slope were estimated with the fixed effect of time. In the second (restricted) model, the fixed effect and random intercept were estimated without the inclusion of a random slope into the model. After both models were estimated, a chi-square difference test was conducted to examine the improvement of fit from the random slope model to the random intercept only model. Estimates of effect size were also calculated for the fixed effect of time and for the model comparison given the large sample sizes across studies.

A summary of the results is presented in Tables I1 and I2. Across both Studies, the estimates of significance and effect size for the fixed effect of time were nearly identical to the results of the RM-ANOVAs. Variance estimates of the random intercepts indicated that there was some individual variability in the starting point on the measures. By contrast, the variance estimates of the random slopes were much smaller, indicating less individual variability in the pattern of measures across the five time occasions. Model comparisons between the random intercept/random slope models and the random intercept only models were all significant but with trivial effect size indicators. This result suggests that the addition of a random slope

[7] Results were nearly identical when setting the covariance structure to be unstructured. We used compound symmetry to be consistent with other reported results.

parameter added little to the models. Specifically, it does not appear that there are meaningful variation patterns among individuals beyond the aggregated fixed effect of time. Thus, the random intercept only models seem to best reflect the data given that these models have better model fit and are more parsimonious.

Table I1

Table of Repeated Measures MLM (RM-MLM) Tests Across Individually-Continuous Time Occasions for Each GAT 1.0 Measure

		Model Random Effects		Model Fit			Test of Fixed Effect of Time (Continuous)				Model Comparisons		
Measure	Model	Intercept s^2	Slope s^2	AIC	BIC	LogLik	b (SE)	p	η^2	90% CI η^2	CM	$\Delta\chi^2$ (Δdf)	Pseudo R^2
Adaptability	M1a: Slp.	0.23	0.01	49,400	49,457	−24,693	−0.00 (0.003)	.761	.000	[.000, .000]	—	—	—
	M1b: Int.	0.24	—	49,643	49,684	−24,817	−0.00 (0.003)	.746	.000	[.000, .000]	M1a	247.41 (2)	.005
Active Coping	M2a: Slp.	0.21	0.01	49,835	49,139	−24,534	−0.01 (0.003)	.036	.000	[.000, .001]	—	—	—
	M2b: Int.	0.26	—	49,468	49,509	−24,729	−0.01 (0.003)	.030	.000	[.000, .001]	M2a	389.73 (2)	.008
Passive Coping	M3a: Slp.	0.34	0.02	63,249	63,306	−31,618	0.05 (0.004)	< .001	.006	[.004, .008]	—	—	—
	M3b: Int.	0.36	—	63,378	63,419	−31,684	0.05 (0.004)	< .001	.007	[.005, .009]	M3a	132.53 (2)	.002
Character	M4a: Slp.	0.86	0.08	84,491	84,548	−42,239	−0.00 (0.007)	.749	.000	[.000, .000]	—	—	—
	M4b: Int.	1.08	—	85,081	85,122	−42,536	−0.00 (0.006)	.989	.000	[.000, .000]	M4a	593.86 (2)	.007
Catastrophizing	M5a: Slp.	0.23	0.01	52,692	52,749	−26,339	−0.03 (0.004)	< .001	.004	[.003, .006]	—	—	—
	M5b: Int.	0.26	—	52,926	52,967	−26,458	−0.03 (0.003)	< .001	.005	[.004, .007]	M5a	237.93 (2)	.005
Depression	M6a: Slp.	0.23	0.01	52,692	52,749	−26,339	−0.03 (0.004)	< .001	.004	[.003, .006]	—	—	—
	M6b: Int.	0.26	—	52,926	52,967	−26,458	−0.03 (0.003)	< .001	.005	[.004, .007]	M6a	237.93 (2)	.005
Optimism	M7a: Slp.	0.36	0.02	51,248	51,305	−25,617	−0.01 (0.004)	.126	.000	[.000, .001]	—	—	—
	M7b: Int.	0.33	—	51,574	51,615	−25,782	−0.01 (0.003)	.081	.000	[.000, .001]	M7a	330.44 (2)	.006
Positive Affect	M8a: Slp.	0.27	0.02	51,943	52,000	−25,965	0.00 (0.004)	.194	.000	[.000, .000]	—	—	—
	M8b: Int.	0.31	—	52,382	52,422	−26,186	0.01 (0.003)	.101	.000	[.000, .001]	M8a	442.47 (2)	.008
Negative Affect	M9a: Slp.	0.20	0.01	46,443	46,500	−23,215	−0.01 (0.003)	< .001	.001	[.000, .001]	—	—	—
	M9b: Int.	0.20	—	46,729	46,770	−23,360	−0.01 (0.003)	< .001	.001	[.000, .002]	M9a	290.12 (2)	.006
Loneliness	M10a: Slp.	0.38	0.02	56,106	56,163	−28,046	−0.01 (0.004)	.001	.001	[.000, .001]	—	—	—
	M10b: Int.	0.39	—	56,372	56,412	−28,181	−0.01 (0.003)	< .001	.001	[.000, .002]	M10a	269.80 (2)	.005
Organizational Trust	M11a: Slp.	0.15	0.02	59,507	59,564	−29,747	−0.08 (0.004)	< .001	.019	[.016, .023]	—	—	—
	M11b: Int.	0.23	—	59,924	59,965	−29,957	−0.08 (0.004)	< .001	.023	[.019, .026]	M11a	420.95 (2)	.007
Work Engagement	M12a: Slp.	0.28	0.03	60,292	60,349	−30,139	−0.09 (0.004)	< .001	.022	[.019, .025]	—	—	—
	M12b: Int.	0.36	—	60,892	60,933	−30,441	−0.09 (0.004)	< .001	.028	[.024, .031]	M12a	603.97 (2)	.010
Life Meaning	M13a: Slp.	0.46	0.03	56,074	56,131	−28,030	0.11 (0.004)	< .001	.035	[.031, .040]	—	—	—
	M13b: Int.	0.38	—	56,409	56,449	−28,199	0.11 (0.003)	< .001	.048	[.043, .053]	M13a	338.88 (2)	.006

Note. GAT = Global Assessment Tool; Model *n*s = 4,966–4,968; *df* = degrees of freedom; Int. = random intercept only model; Slp. = random intercept and slope model; AIC = Akaike information criterion; BIC = Bayesian information criterion; LogLik = log likelihood; CM = comparison model; Δ = change; All $\Delta\chi^2$ *p*s < .001; pseudo R^2 calculated using McFadden approximation; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988).

Table I2

Table of Repeated Measures MLM (RM-MLM) Tests Across Individually-Continuous Time Occasions for Each GAT 2.0 Measure

		Model Random Effects		Model Fit			Test of Fixed Effect of Time (Continuous)				Model Comparisons		
Measure	Model	Intercept s^2	Slope s^2	AIC	BIC	LogLik	*b* (*SE*)	*p*	η^2	90% CI η^2	CM	$\Delta\chi^2$ (Δdf)	Pseudo R^2
Adaptability	M1a: Slp.	0.31	0.01	6,414	6,457	−3,200	−0.03 (0.009)	< .001	.004	[.001, .009]	—	—	—
	M1b: Int.	0.32	—	6,428	6,459	−3,209	−0.03 (0.010)	.002	.004	[.001, .010]	M1a	17.70 (2)	.003
Active Coping	M2a: Slp.	0.30	0.01	6,196	6,238	−3,091	−0.04 (0.010)	< .001	.006	[.002, .011]	—	—	—
	M2b: Int.	0.31	—	6,218	6,248	−3,104	−0.04 (0.009)	< .001	.007	[.003, .014]	M2a	26.24 (2)	.004
Passive Coping	M3a: Slp.	0.33	0.02	8,140	8,182	−4,063	0.03 (0.014)	.023	.002	[.000, .006]	—	—	—
	M3b: Int.	0.32	—	8,147	8,177	−4,068	0.03 (0.013)	.013	.002	[.000, .001]	M3a	10.97 (2)	.001
Character	M4a: Slp.	1.04	0.09	10,408	10,450	−5,197	−0.10 (0.020)	< .001	.010	[.004, .017]	—	—	—
	M4b: Int.	1.29	—	10,490	10,521	−5,240	−0.10 (0.017)	< .001	.012	[.006, .020]	M4a	86.73 (2)	.008
Catastrophizing	M5a: Slp.	0.37	0.02	7,733	7,776	−3,859	0.05 (0.013)	< .001	.007	[.003, .013]	—	—	—
	M5b: Int.	0.41	—	7,756	7,787	−3,873	0.05 (0.011)	< .001	.008	[.003, .014]	M5a	27.10 (2)	.004
Depression	M6a: Slp.	0.22	0.03	6,170	6,212	−3,078	0.05 (0.011)	< .001	.009	[.004, .016]	—	—	—
	M6b: Int.	0.25	—	6,258	6,289	−3,124	0.05 (0.009)	< .001	.012	[.006, .020]	M6a	92.37 (2)	.015
Optimism	M7a: Slp.	0.39	0.06	1,485	1,517	−736	−0.02 (0.033)	.550	.001	[.000, .009]	—	—	—
	M7b: Int.	0.37	—	1,498	1,521	−744	−0.02 (0.027)	.485	.001	[.000, .010]	M7a	16.61 (2)	.011
Positive Affect	M8a: Slp.	0.38	0.06	1,475	1,507	−731	−0.04 (0.033)	.186	.003	[.000, .015]	—	—	—
	M8b: Int.	0.43	—	1,490	1,513	−740	−0.04 (0.027)	.095	.005	[.000, .019]	M8a	19.31 (2)	.013
Negative Affect	M9a: Slp.	0.22	0.04	1,317	1,349	−651	0.05 (0.030)	.101	.005	[.000, .019]	—	—	—
	M9b: Int.	0.24	—	1,327	1,350	−659	0.05 (0.025)	.065	.006	[.000, .021]	M9a	14.14 (2)	.011
Loneliness	M10a: Slp.	0.45	0.03	7,245	7,288	−3,616	0.01 (0.012)	.299	.000	[.000, .003]	—	—	—
	M10b: Int.	0.43	—	7,297	7,328	−3,644	0.01 (0.011)	.222	.001	[.000, .003]	M10a	56.06 (2)	.008
Organizational Trust	M11a: Slp.	0.26	0.04	7,524	7,567	−3,755	−0.13 (0.014)	< .001	.034	[.024, .047]	—	—	—
	M11b: Int.	0.29	—	7,594	7,624	−3,792	−0.13 (0.011)	< .001	.044	[.032, .057]	M11a	73.30 (2)	.010
Work Engagement	M12a: Slp.	0.43	0.07	7,826	7,869	−3,906	−0.13 (0.015)	< .001	.026	[.017, .036]	—	—	—
	M12b: Int.	0.45	—	7,956	7,987	−3,973	−0.12 (0.012)	< .001	.036	[.025, .048]	M12a	133.60 (2)	.017
Life Meaning	M13a: Slp.	0.31	0.03	6,196	6,212	−3,077	−0.06 (0.011)	< .001	.011	[.006, .019]	—	—	—
	M13b: Int.	0.35	—	6,257	6,288	−3,124	−0.06 (0.009)	< .001	.016	[.009, .024]	M13a	92.47 (2)	.015

Note. GAT = Global Assessment Tool; Model *n*s = 142–658; *df* = degrees of freedom; Int. = random intercept only model; Slp. = random intercept and slope model; AIC = Akaike information criterion; BIC = Bayesian information criterion; LogLik = log likelihood; CM = comparison model; Δ = change; All $\Delta\chi^2$ *p*s < .01; pseudo R^2 calculated using McFadden approximation; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988).

Appendix J: SEM Analysis of Random Effects: Intercepts and Slopes

In addition to the fixed effects reported in the main text, we also examined the random effects for the scale-level models using SEM (see Figure A7). Specifically, we obtained estimates of the random intercepts and random slopes for each of the GAT measures. Estimates of random intercepts indicate the degree to which there is individual variability in the starting point for each respondent (i.e., value at time zero). Estimates of random slopes indicate the degree to which there is individual variability in the slope across time (i.e., the pattern in which values change over time).

To examine these estimates, we conducted two nested growth curve models using the lavaan SEM package in R (Little, 2013; Rosseel, 2014) for the GAT 1.0 data in Study 1 and the GAT 2.0 data in Study 2. For these models, the scale-level composites (aggregated mean across individual items) of the thirteen measures were used in separate tests. Five time occasions (coded 0, 1, 2, 3, 4, respectively) were modeled for each measure. Models were estimated with maximum likelihood (ML) and missing data were treated with full-information maximum likelihood (FIML). In the first (full) model, the random intercept and random slope were estimated. In the second (restricted) model, the random intercept was estimated but the random slope was fixed to zero (for detailed code of model parameters, see Appendix J). After both models were estimated, a chi-square difference test was conducted to examine the improvement of fit from the random slope model to the intercept only model. Estimates of effect size were also calculated given the large sample sizes across studies.

A summary of the results is presented in Tables H1 and H2. Across both Studies, the variance estimates for random intercepts and slopes were significant. This result is not surprising given the large sample sizes of the two studies. Variance estimates of the random intercepts indicated that there was some individual variability in the starting point on the measures. By contrast, the variance estimates of the random slopes were much smaller, indicating less individual variability in the pattern of measures across the five time occasions. Lastly, the results of the nested model comparison tests indicated that, for many of the measures, there was a significant increase in model fit when considering models with both a random intercept and slope (vs. random intercept only). However, effect size indicators for these comparisons were trivial to small (especially in Study 2). Consistent with the findings presented in the main text, measures that did show a small effect tended to be the same measures that showed small fixed effects over time (e.g., life meaning, organizational trust, work engagement). Of note, the character measure did show the highest degree of random intercept and random slope variability across studies.

Table J1

Table of Random Intercept and Random Slope SEM Tests Across Five Time Occasions for Each GAT 1.0 Measure

		Model Random Effects		Model Fit Statistics					Model Comparisons					Measures of Effect Size			
Measure	Model	Intercept s^2 (*SE* of Est.)	Slope s^2 (*SE* of Est.)	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2	
Adaptability	M1a: Slp.	0.24 (0.002)	0.01 (0.001)	704.97 (14)	.985	.989	.023	.072	—	—	—	—	—	—	—	—	
	M1b: Int.	0.25 (0.002)	FS	1,688.72 (16)	.964	.977	.033	.061	M1a	983.75 (2)	−.021	−.012	.010	0.072	.011	[.010, .012]	
Active Coping	M2a: Slp.	0.22 (0.002)	0.01 (0.001)	1,119.03 (14)	.977	.984	.029	.098	—	—	—	—	—	—	—	—	
	M2b: Int.	0.26 (0.002)	FS	3,492.18 (16)	.929	.956	.048	.097	M2a	2,373.14 (2)	−.048	−.028	.019	0.112	.026	[.025, .027]	
Passive Coping	M3a: Slp.	0.32 (0.004)	0.01 (0.001)	459.03 (14)	.987	.991	.018	.043	—	—	—	—	—	—	—	—	
	M3b: Int.	0.34 (0.003)	FS	1,016.37 (16)	.972	.982	.026	.051	M3a	557.34 (2)	−.015	−.009	.008	0.054	.006	[.006, .007]	
Character	M4a: Slp.	0.81 (0.009)	0.08 (0.003)	1,844.04 (14)	.964	.974	.037	.128	—	—	—	—	—	—	—	—	
	M4b: Int.	1.06 (0.007)	FS	6,566.51 (16)	.870	.919	.066	.135	M4a	4,722.47 (2)	−.094	−.055	.029	0.157	.051	[.049, .052]	
Catastrophizing	M5a: Slp.	0.21 (0.002)	0.01 (0.001)	865.52 (14)	.980	.985	.025	.073	—	—	—	—	—	—	—	—	
	M5b: Int.	0.25 (0.002)	FS	2,504.27 (16)	.940	.963	.040	.085	M5a	1,638.74 (2)	−.040	−.022	.015	0.093	.018	[.017, .019]	
Depression	M6a: Slp.	0.19 (0.002)	0.01 (0.001)	1,512.10 (14)	.958	.970	.034	.107	—	—	—	—	—	—	—	—	
	M6b: Int.	0.23 (0.002)	FS	3,602.06 (16)	.899	.937	.049	.095	M6a	2,089.96 (2)	−.059	−.033	.015	0.105	.023	[.022, .024]	
Optimism	M7a: Slp.	0.31 (0.003)	0.02 (0.001)	1,073.19 (14)	.982	.987	.028	.061	—	—	—	—	—	—	—	—	
	M7b: Int.	0.31 (0.002)	FS	2,384.76 (16)	.960	.975	.041	.068	M7a	1,311.57 (2)	−.022	−.012	.011	0.083	.015	[.014, .015]	
Positive Affect	M8a: Slp.	0.27 (0.003)	0.02 (0.001)	1,285.70 (14)	.977	.983	.031	.090	—	—	—	—	—	—	—	—	
	M8b: Int.	0.31 (0.002)	FS	3,550.42 (16)	.935	.959	.048	.095	M8a	2,264.72 (2)	−.042	−.024	.017	0.109	.025	[.024, .026]	
Negative Affect	M9a: Slp.	0.19 (0.002)	0.01 (0.001)	1,077.66 (14)	.975	.982	.028	.085	—	—	—	—	—	—	—	—	
	M9b: Int.	0.21 (0.002)	FS	2,541.12 (16)	.941	.963	.041	.077	M9a	1,463.46 (2)	−.034	−.019	.013	0.088	.016	[.015, .017]	
Loneliness	M10a: Slp.	0.34 (0.003)	0.02 (0.001)	789.89 (14)	.986	.990	.024	.072	—	—	—	—	—	—	—	—	
	M10b: Int.	0.37 (0.002)	FS	2,217.28 (16)	.961	.976	.038	.072	M10a	1,427.39 (2)	−.025	−.014	.014	0.087	.016	[.015, .017]	
Organizational Trust	M11a: Slp.	0.14 (0.003)	0.02 (0.001)	3,100.57 (14)	.877	.912	.048	.142	—	—	—	—	—	—	—	—	
	M11b: Int.	0.23 (0.002)	FS	7,151.64 (16)	.717	.823	.068	.133	M11a	4,051.07 (2)	−.160	−.089	.020	0.146	.044	[.042, .045]	
Work Engagement	M12a: Slp.	0.30 (0.004)	0.03 (0.001)	4,938.29 (14)	.886	.919	.061	.173	—	—	—	—	—	—	—	—	
	M12b: Int.	0.38 (0.003)	FS	8,738.50 (16)	.798	.874	.076	.135	M12a	3,800.21 (2)	−.088	−.045	.015	0.141	.041	[.040, .043]	
Life Meaning	M13a: Slp.	0.40 (0.003)	0.02 (0.001)	2,757.23 (14)	.954	.967	.045	.105	—	—	—	—	—	—	—	—	
	M13b: Int.	0.37 (0.002)	FS	3,731.07 (16)	.937	.961	.049	.075	M13a	973.84 (2)	−.017	−.006	.004	0.071	.011	[.010, .012]	

Note. GAT = Global Assessment Tool; Model *n*s = 95,277; All random effect estimates *p*s < .001; All $\Delta\chi^2$ *p*s < .001; *df* = degrees of freedom; Int. = random intercept with fixed slope; Slp. = random intercept and slope; FS = fixed slope to 0; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Table J2

Table of Random Intercept and Random Slope SEM Tests Across Five Time Occasions for Each GAT 2.0 Measure

		Model Random Effects		Model Fit Statistics					Model Comparisons					Measures of Effect Size		
Measure	Model	Intercept s^2 (*SE* of Est.)	Slope s^2 (*SE* of Est.)	χ^2 (*df*)	CFI	TLI	RMSEA	SRMR	CM	$\Delta\chi^2$ (Δdf)	ΔCFI	ΔTLI	ΔRMSEA	ω	η^2	90% CI η^2
Adaptability	M1a: Slp.	0.24 (0.003)	0.01 (0.001)	276.98 (14)	.988	.992	.018	.104	—	—	—	—	—	—	—	—
	M1b: Int.	0.26 (0.002)	FS	679.36 (16)	.970	.981	.027	.069	M1a	402.38 (2)	−.018	−.011	.009	0.046	.005	[.004, .005]
Active Coping	M2a: Slp.	0.23 (0.003)	0.01 (0.001)	407.52 (14)	.984	.988	.022	.131	—	—	—	—	—	—	—	—
	M2b: Int.	0.26 (0.002)	FS	1,253.94 (16)	.948	.968	.037	.090	M2a	846.42 (2)	−.036	−.020	.015	0.067	.009	[.009, .010]
Passive Coping	M3a: Slp.	0.30 (0.005)	0.01 (0.002)	87.10 (14)	.995	.996	.010	.050	—	—	—	—	—	—	—	—
	M3b: Int.	0.31 (0.003)	FS	202.20 (16)	.987	.992	.014	.041	M3a	115.10 (2)	−.008	−.004	.004	0.025	.001	[.001, .002]
Character	M4a: Slp.	0.79 (0.011)	0.07 (0.005)	844.37 (14)	.968	.977	.032	.205	—	—	—	—	—	—	—	—
	M4b: Int.	1.06 (0.009)	FS	3,316.79 (16)	.875	.922	.060	.146	M4a	2,472.42 (2)	−.093	−.055	.028	0.114	.027	[.026, .028]
Catastrophizing	M5a: Slp.	0.28 (0.005)	0.01 (0.002)	555.12 (14)	.967	.976	.026	.113	—	—	—	—	—	—	—	—
	M5b: Int.	0.34 (0.003)	FS	1,182.93 (16)	.929	.956	.036	.082	M5a	627.81 (2)	−.038	−.020	.010	0.057	.007	[.006, .008]
Depression	M6a: Slp.	0.15 (0.003)	0.02 (0.002)	839.63 (14)	.950	.964	.032	.224	—	—	—	—	—	—	—	—
	M6b: Int.	0.21 (0.002)	FS	2,803.08 (16)	.830	.894	.055	.139	M6a	1,963.44 (2)	−.120	−.070	.023	0.102	.022	[.021, .023]
Optimism	M7a: Slp.	0.28 (0.004)	0.02 (0.002)	119.29 (14)	.994	.995	.011	.163	—	—	—	—	—	—	—	—
	M7b: Int.	0.29 (0.003)	FS	567.81 (16)	.967	.979	.024	.111	M7a	448.52 (2)	−.027	−.016	.013	0.049	.005	[.004, .006]
Positive Affect	M8a: Slp.	0.27 (0.004)	0.02 (0.003)	202.92 (14)	.987	.991	.015	.212	—	—	—	—	—	—	—	—
	M8b: Int.	0.30 (0.003)	FS	968.26(16)	.936	.960	.032	.129	M8a	765.34 (2)	−.051	−.031	.017	0.063	.009	[.008, .009]
Negative Affect	M9a: Slp.	0.21 (0.003)	0.02 (0.002)	147.06 (14)	.988	.991	.013	.145	—	—	—	—	—	—	—	—
	M9b: Int.	0.21 (0.002)	FS	450.89 (16)	.961	.975	.022	.087	M9a	303.83 (2)	−.027	−.016	.009	0.040	.003	[.003, .004]
Loneliness	M10a: Slp.	0.34 (0.004)	0.02 (0.002)	294.64 (14)	.990	.993	.019	.116	—	—	—	—	—	—	—	—
	M10b: Int.	0.37 (0.003)	FS	884.88 (16)	.968	.980	.031	.076	M10a	590.24 (2)	−.022	−.013	.012	0.056	.007	[.006, .007]
Organizational Trust	M11a: Slp.	0.15 (0.004)	0.02 (0.002)	1,275.77 (14)	.897	.926	.039	.257	—	—	—	—	—	—	—	—
	M11b: Int.	0.24 (0.003)	FS	3,555.57 (16)	.711	.819	.062	.144	M11a	2,279.80 (2)	−.186	−.107	.023	0.109	.025	[.024, .026]
Work Engagement	M12a: Slp.	0.29 (0.005)	0.04 (0.002)	1,570.74 (14)	.925	.947	.044	.270	—	—	—	—	—	—	—	—
	M12b: Int.	0.37 (0.004)	FS	3,673.59 (16)	.824	.890	.063	.156	M12a	2,102.85 (2)	−.101	−.057	.019	0.105	.023	[.022, .024]
Life Meaning	M13a: Slp.	0.22 (0.003)	0.02 (0.001)	829.62 (14)	.970	.979	.032	.220	—	—	—	—	—	—	—	—
	M13b: Int.	0.29 (0.003)	FS	3,121.42 (16)	.886	.929	.058	.143	M13a	2,291.80 (2)	−.084	−.050	.026	0.110	.025	[.024, .026]

Note. GAT = Global Assessment Tool; Model *n*s = 57,495–57,771; All random effect estimates *p*s < .001; All $\Delta\chi^2$ *p*s < .001. *df* = degrees of freedom; Int. = random intercept with fixed slope; Slp. = random intercept and slope; FS = fixed slope to 0; CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square error of approximation; CM = comparison model; Δ = change; Guidelines for η^2 effect size interpretation: .02 = small, .13 = medium, .26 = large (Cohen, 1988); ω = Cohen's ω; Guidelines for ω effect size interpretation: 0.10 = small, 0.30 = medium, 0.50 = large (Cohen, 1988).

Appendix K: Mean-Level Change, Reliable Change, and Rank-Order Stability

We also assessed the following forms of longitudinal stability for each of the 13 measures across two time occasions: mean-level change, reliable change, and rank-order stability (see Roberts et al., 2008). Each of these methods is described in some detail below.

Mean-Level Change

Mean-level change examines the stability of measures by assessing the levels of measurement change that occur *on average* with time. Here, we assess mean-level change with paired-sample *t*-tests between Time Occasion 1 and Time Occasion 2 as well as Time Occasion 1 and Time Occasion 5. Cohen's *d* was used as an effect size indicator for the magnitude of change that occurs between these time occasions. Cohen's *d* has the following suggested guidelines for interpretation: small = 0.20, medium = 0.50, large = 0.80 (see Cohen, 1988).

Reliable Change

An extension of mean-level change and common in the personality literature (see Robins et al., 2001), reliable change examines absolute change at the individual level. Individual differences in change are calculated using a Reliable Change (RC; see Christensen & Mendoza, 1986; De Fruyt et al., 2006; Ibáñez et al., 2016; Jacobson & Traux, 1991; Robbins et al., 2001). The RC calculates the probability of observing a difference score equal to, or greater than that obtained with a no change null hypothesis. Importantly, the index accounts for measurement error by incorporating measure reliability (e.g., test-retest), which parses out whether increases or decreases in scores go above and beyond what could be accounted for by measurement error. The following formula is used to calculate RC: $RC = \frac{X_2 - X_1}{S_{diff}}$, where X_1 represents a person's score at Time Occasion 1, X_2 represents a person's score at Time Occasion 2, and S_{diff} is the standard error of difference between the two test scores. The formula for is: $S_{\text{diff}} = \sqrt{2\ (SE_{M^2})}$, where SE_M is the standard error of measurement which is defined as $SE_{\text{M}} = s\sqrt{1 - r_{xx}}$ (r_{xx} = measure reliability such as omega total or test-retest; *s* = measure standard deviation at Time Occasion 1 as a reference group). RC standardizes the difference in *SD* units and if RC > 1.96 or < −1.96, then the change is considered significantly reliable beyond measurement error. Another metric, the reliable change index (RCI) keeps the measure in its original units and provides the absolute value of the difference score required for RC to be reliable. RCI is calculated using the following formula: $RCI = S_{\text{diff}} \times 1.96$; if the absolute value of the difference score between time points is greater than this RCI value, then there is reliable change. For more details on RC and RCI, see De Fruyt et al. (2006).

Computing RCs for each measure enables the determination of how many individuals remain stable (i.e., indeterminate change or RC0%) on the measures across time occasions but also provides information on the frequency of individual-level change patterns on the measures in terms of increases (RC+%) and decreases RC−%). The combined percentage of RC+% plus RC−% would give the magnitude for any reliable change (RC±%). RC can be thought of as similar (albeit more conservative) to the common language effect size (CLES) or probability of superiority (Blampied, 2017; Lakens, 2013; McGraw & Wong, 1992); individual-level RC

indicates the magnitude of the increase or decrease in any given measure about a person over time. We classified an individual as having increased or decreased on each measure if the probability associated with their RCI score was less than 5%. Suggested guidelines for interpretation for the magnitude of percentages include small (56%), medium (64%), and large (72%) when examining if any change occurred (i.e., RC±%) and small (11%), medium (28%), and large (43%) when looking at either RC+% or RC−% (Vargha & Delaney, 2000).

Rank-Order Stability

Rank-order stability assesses whether the rank order of individuals in a certain measure is maintained over time. We assessed rank-order stability by calculating the correlation of a measure between time occasions (i.e., Time Occasion 1 vs. Time Occasion 2, Time Occasion 1 vs. Time Occasion 5).

Results Summary

A summary of results is presented in Tables K1 through K4.

Results of *mean-level change* indicated that most measures did not show a magnitude of change that reached the small effect threshold of 0.20 as indicated by Cohen's *d* estimates. However, as indicated by the other methods presented in this report, organizational trust, work engagement, and life meaning did show small effects. These findings held consistent across GAT versions and whether comparing Time Occasion 1 with 2 or 5.

Results of the *individual-level reliable change* using RC indicated that many of the measures showed high frequencies of indeterminate change, suggesting stability in the measurement across time. However, measures like character, depression, positive affect, and negative affect did show moderate increases and decreases across GAT versions and whether comparing Time Occasion 1 with 2 or 5. This suggests that for some measures, individual-level change was occurring above and beyond what could be accountable to measurement error.

Results of *rank-order stability* indicated that most measures showed a medium to large correlation across time occasions. Measures like optimism, positive affect, life meaning, and loneliness had high correlations between time occasions. By contrast, organizational trust and work engagement had relatively lower correlations between time occasions, albeit still moderate in magnitude.

Table K1

Mean-Level Change, Reliable Change, and Rank-Order Stability for GAT 1.0 (Time 1 vs. Time 2)

Measure	*n*	Time Occasion 1 *M* (*SD*)	Time Occasion 2 *M* (*SD*)	*Mean Change (Cohen's d)* [a]	90% CI *d*	Individual-Level Change (Reliable Change) [b]			Rank-Order Stability [c]
						Reliable Decrease	Indeterminate Change	Reliable Increase	
Adaptability	95,258	4.06 (0.74)	4.05 (0.78)	(−) 0.013*	[0.000, 0.026]	8%	85%	7%	,42 (.58)
Active Coping	95,256	3.85 (0.71)	3.82 (0.79)	(−) 0.034*	[0.021, 0.046]	12%	77%	11%	.43 (.55)
Passive Coping	95,256	2.60 (0.92)	2.68 (0.96)	(+) 0.077*	[0.064, 0.089]	6%	85%	8%	.37 (.51)
Character	95,254	7.98 (1.37)	7.97 (1.60)	(−) 0.004	[0.009, 0.017]	23%	53%	24%	.43 (.45)
Catastrophizing	95,260	1.93 (0.73)	1.88 (0.80)	(−) 0.060*	[0.048, 0.073]	15%	73%	13%	.39 (.46)
Depression	95,263	1.60 (0.72)	1.64 (0.81)	(+) 0.049*	[0.037, 0.062]	16%	65%	19%	.36 (.46)
Optimism	95,259	3.71 (0.79)	3.68 (0.80)	(−) 0.046*	[0.033, 0.059]	8%	84%	7%	.47 (.64)
Positive Affect	95,269	3.82 (0.77)	3.80 (0.84)	(−) 0.025*	[0.012, 0.038]	20%	62%	18%	.45 (.48)
Negative Affect	95,269	2.20 (0.68)	2.19 (0.73)	(−) 0.022*	[0.009, 0.034]	16%	68%	15%	.40 (.45)
Loneliness	95,259	2.21 (0.85)	2.21 (0.90)	(−) 0.003	[0.009, 0.016]	9%	82%	9%	.47 (.58)
Organizational Trust	95,262	3.98 (0.75)	3.81 (0.90)	(−) 0.172*	[0.160, 0.185]	20%	68%	12%	.31 (.37)
Work Engagement	95,259	3.87 (0.88)	3.59 (1.01)	(−) 0.260*	[0.248, 0.273]	19%	74%	7%	.39 (.48)
Life Meaning	95,256	3.87 (0.86)	3.91 (0.92)	(+) 0.048*	[0.035, 0.060]	9%	81%	10%	.50 (.61)

Note. * $p < .05$.

[a] Mean-level change in standard score units (i.e., Cohen's *d*). Absolute value is reported. Positive and negative signs in parentheses indicate direction of change.

[b] Percentage of individuals who decreased, had stable scores, or increased on each measure, according to reliable change (see text). The expected frequencies are 2.5%, 95%, and 2.5%, respectively.

[C] Correlations between Time Occasion 1 and Time Occasion 2. Values in parentheses are disattenuated for measurement error (based on Omega Total reliability).

Table K2

Mean-Level Change, Reliable Change, and Rank-Order Stability for GAT 1.0 (Time 1 vs. Time 5)

Measure	*n*	Time Occasion 1 *M* (*SD*)	Time Occasion 5 *M* (*SD*)	*Mean Change (Cohen's d)* [a]	90% CI *d*	Individual-Level Change (Reliable Change) [b]			Rank-Order Stability [c]
						Reliable Decrease	Indeterminate Change	Reliable Increase	
Adaptability	4,967	4.08 (0.71)	4.09 (0.77)	(+) 0.008	[0.048, 0.064]	9%	84%	7%	.35 (.48)
Active Coping	4,967	3.88 (0.67)	3.87 (0.80)	(−) 0.021	[−0.034, 0.077]	13%	76%	11%	.36 (.45)
Passive Coping	4,967	2.62 (0.90)	2.80 (1.00)	(+) 0.161*	[0.105, 0.217]	6%	83%	11%	.32 (.44)
Character	4,966	8.00 (1.33)	8.01 (1.69)	(+) 0.011	[−0.045, 0.066]	24%	48%	28%	.34 (.35)
Catastrophizing	4,967	1.96 (0.72)	1.84 (0.84)	(−) 0.143*	[0.088, 0.199]	19%	68%	12%	.33 (.38)
Depression	4,967	1.59 (0.72)	1.63 (0.81)	(+) 0.044*	[−0.012, 0.099]	18%	62%	20%	.27 (.29)
Optimism	4,967	3.76 (0.83)	3.74 (0.83)	(−) 0.021	[−0.077, 0.034]	10%	81%	9%	.41 (.55)
Positive Affect	4,968	3.83 (0.75)	3.86 (0.87)	(+) 0.032*	[−0.024, 0.088]	20%	57%	22%	.38 (.40)
Negative Affect	4,968	2.19 (0.67)	2.14 (0.74)	(−) 0.065*	[0.009, 0.121]	19%	65%	15%	.32 (.36)
Loneliness	4,967	2.24 (0.84)	2.19 (0.92)	(−) 0.050*	[−0.005, 0.106]	10%	80%	10%	.42 (.51)
Organizational Trust	4,968	3.98 (0.75)	3.72 (0.95)	(−) 0.239*	[0.183, 0.295]	25%	62%	12%	.20 (.24)
Work Engagement	4,967	3.98 (0.80)	3.65 (1.00)	(−) 0.307*	[0.251, 0.363]	21%	72%	7%	.29 (.35)
Life Meaning	4,967	3.85 (0.84)	4.15 (0.86)	(+) 0.309*	[0.253, 0.365]	7%	75%	18%	.36 (.43)

Note. * $p < .05$.

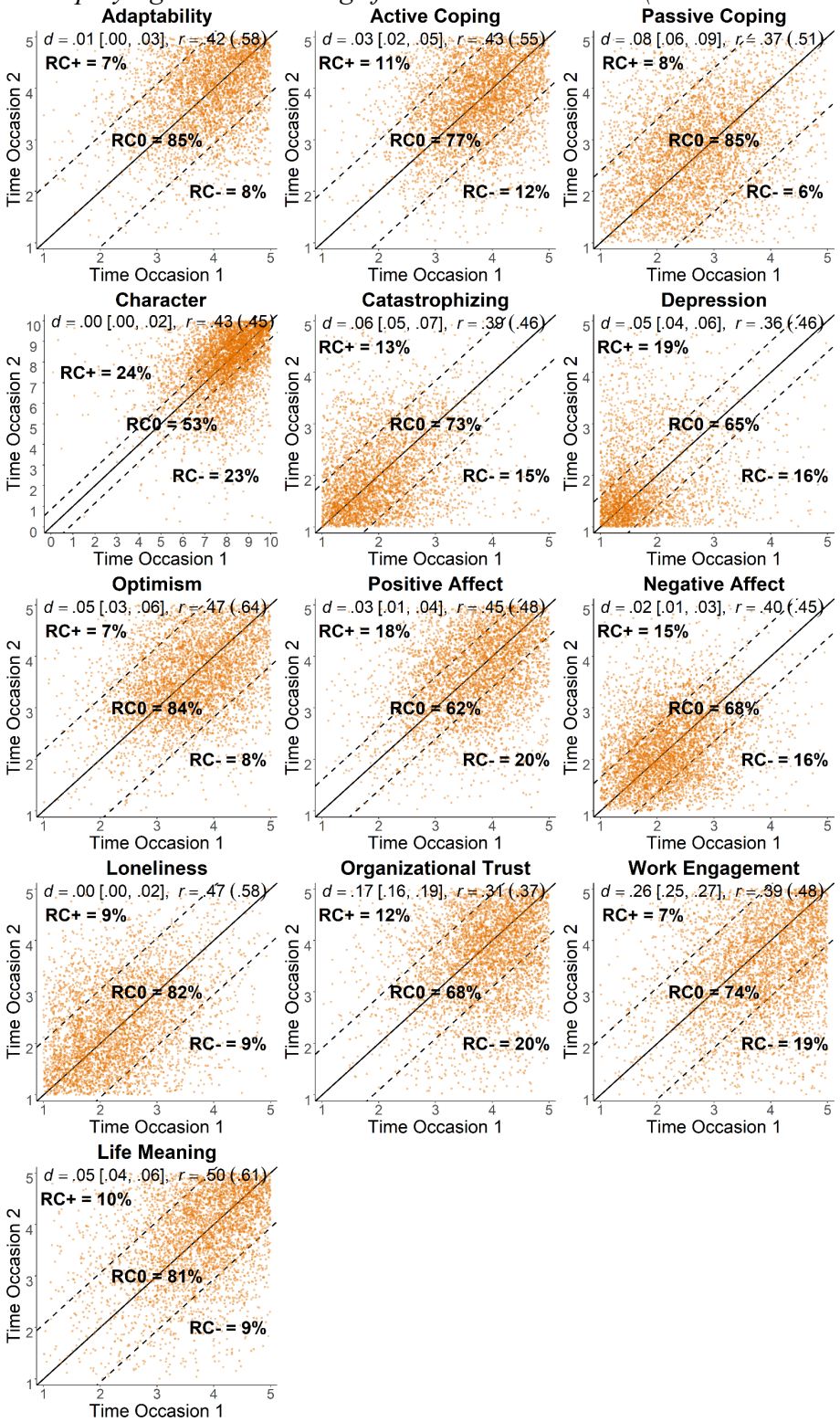
[a] Mean-level change in standard score units (i.e., Cohen's *d*). Absolute value is reported. Positive and negative signs in parentheses indicate direction of change.

[b] Percentage of individuals who decreased, had stable scores, or increased on each measure, according to reliable change (see text). The expected frequencies are 2.5%, 95%, and 2.5%, respectively.

[C] Correlations between Time Occasion 1 and Time Occasion 5. Values in parentheses are disattenuated for measurement error (based on Omega Total reliability).

Figure K1

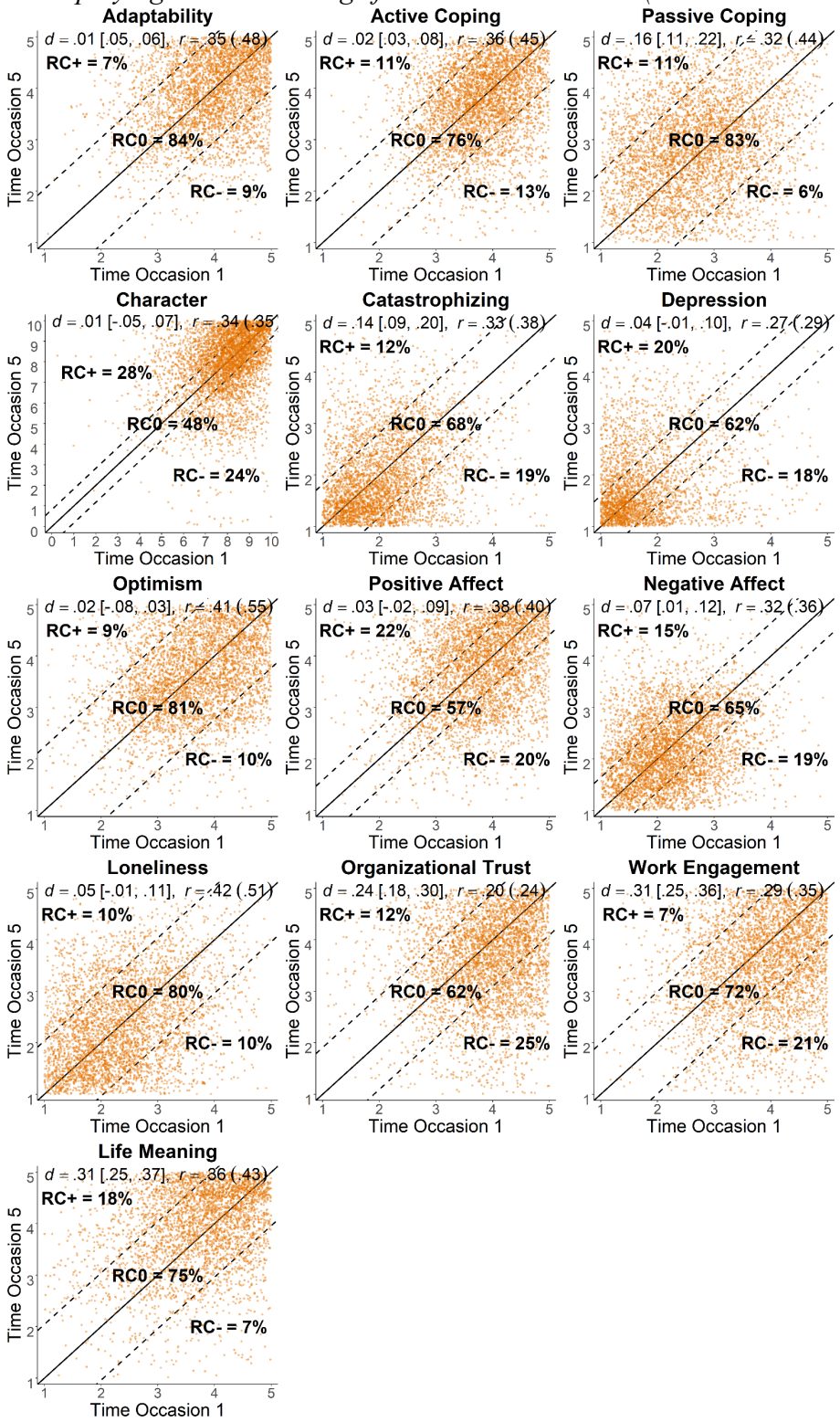
Modified Brinley Plots Displaying Reliable Change for GAT 1.0 Measures (Time 1 vs. Time 2)



Note. Modified Brinley plots showing individual-level change between Time Occasion 1 and Time Occasion 2. Points represent random subsample of 5,000 respondents (*n*s = 95,254–95,269). Diagonal solid lines = no change; diagonal dotted lines = reliable change (RC) upper and lower boundaries; RC+ = Reliable Change increase; RC− = Reliable Change decrease; RC0 = Indeterminate Change. *d* = Cohen's *d* with 95% confidence interval for mean-level change. *r* = rank-order stability correlation with attenuated correlation coefficient in parentheses.

Figure K2

Modified Brinley Plots Displaying Reliable Change for GAT 1.0 Measures (Time 1 vs. Time 5)



Note. Modified Brinley plots showing individual-level change between Time Occasion 1 and Time Occasion 5. Points represent sample of respondents (*n*s = 4,967–4,968). Diagonal solid lines = no change; diagonal dotted lines = reliable change (RC) upper and lower boundaries; RC+ = Reliable Change increase; RC− = Reliable Change decrease; RC0 = Indeterminate Change. *d* = Cohen's *d* with 95% confidence interval for mean-level change. *r* = rank-order stability correlation with attenuated correlation coefficient in parentheses.

Table K3

Mean-Level Change, Reliable Change, and Rank-Order Stability for GAT 2.0 (Time 1 vs. Time 2)

Measure	*n*	Time Occasion 1 *M* (*SD*)	Time Occasion 2 *M* (*SD*)	*Mean Change (Cohen's d)* [a]	90% CI *d*	Individual-Level Change (Reliable Change) [b]			Rank-Order Stability [c]
						Reliable Decrease	Indeterminate Change	Reliable Increase	
Adaptability	57,720	4.09 (0.73)	4.02 (0.77)	(−) 0.092*	[0.076, 0.108]	9%	86%	6%	.45 (.65)
Active Coping	57,719	3.94 (0.70)	3.88 (0.78)	(−) 0.078*	[0.062, 0.094]	12%	79%	9%	.47 (.59)
Passive Coping	57,719	2.50 (0.90)	2.52 (0.92)	(+) 0.017*	[0.000, 0.033]	7%	87%	7%	.37 (.52)
Character	57,718	8.18 (1.32)	8.08 (1.57)	(−) 0.071*	[0.055, 0.087]	22%	59%	19%	.49 (.52)
Catastrophizing	57,718	1.92 (0.87)	2.03 (0.96)	(+) 0.114*	[0.098, 0.130]	8%	80%	12%	.40 (.53)
Depression	57,715	1.51 (0.67)	1.60 (0.78)	(+) 0.115*	[0.099, 0.131]	13%	69%	19%	.39 (.42)
Optimism	35,910	3.72 (0.73)	3.66 (0.78)	(−) 0.080*	[0.059, 0.100]	7%	88%	5%	.53 (.72)
Positive Affect	35,903	3.95 (0.75)	3.87 (0.83)	(−) 0.100*	[0.080, 0.121]	21%	63%	16%	.49 (.52)
Negative Affect	35,903	2.02 (0.68)	2.04 (0.73)	(+) 0.028*	[0.008, 0.049]	16%	66%	18%	.44 (.49)
Loneliness	57,717	2.21 (0.82)	2.26 (0.88)	(+) 0.059*	[0.043, 0.076]	6%	84%	9%	.50 (.63)
Organizational Trust	57,720	4.10 (0.75)	3.90 (0.89)	(−) 0.210*	[0.194, 0.227]	22%	67%	11%	.34 (.40)
Work Engagement	57,720	3.96 (0.84)	3.67 (0.98)	(−) 0.288*	[0.271, 0.304]	17%	77%	6%	.43 (.52)
Life Meaning	57,719	4.32 (0.69)	4.19 (0.80)	(−) 0.167*	[0.151, 0.183]	14%	79%	7%	.50 (.60)

Note. * $p < .05$.

[a] Mean-level change in standard score units (i.e., Cohen's *d*). Absolute value is reported. Positive and negative signs in parentheses indicate direction of change.

[b] Percentage of individuals who decreased, had stable scores, or increased on each measure, according to reliable change (see text). The expected frequencies are 2.5%, 95%, and 2.5%, respectively.

[C] Correlations between Time Occasion 1 and Time Occasion 2. Values in parentheses are disattenuated for measurement error (based on Omega Total reliability).

Table K4

Mean-Level Change, Reliable Change, and Rank-Order Stability for GAT 2.0 (Time 1 vs. Time 5)

Measure	*n*	Time Occasion 1 *M* (*SD*)	Time Occasion 5 *M* (*SD*)	*Mean Change (Cohen's d)* [a]	90% CI *d*	Individual-Level Change (Reliable Change) [b]			Rank-Order Stability [c]
						Reliable Decrease	Indeterminate Change	Reliable Increase	
Adaptability	667	4.10 (0.73)	4.01 (0.80)	(−) 0.114*	[−0.038, 0.266]	8%	85%	6%	.45 (.63)
Active Coping	667	3.98 (0.70)	3.86 (0.81)	(−) 0.152*	[0.000, 0.304]	14%	78%	8%	.48 (.60)
Passive Coping	667	2.47 (0.87)	2.59 (0.95)	(+) 0.117*	[−0.269, 0.035]	5%	86%	9%	.34 (.47)
Character	667	8.30 (1.25)	8.02 (1.66)	(−) 0.187*	[0.035, 0.339]	27%	55%	18%	.49 (.51)
Catastrophizing	667	1.92 (0.85)	2.07 (1.00)	(+) 0.150*	[−0.002, 0.302]	7%	80%	14%	.41 (.53)
Depression	666	1.48 (0.63)	1.62 (0.80)	(+) 0.166*	[0.014, 0.318]	21%	67%	12%	.33 (.36)
Optimism	146	3.79 (0.81)	3.73 (0.92)	(−) 0.074	[−0.251, 0.398]	10%	84%	6%	.49 (.63)
Positive Affect	146	3.93 (0.79)	3.92 (0.93)	(−) 0.008	[−0.317, 0.332]	25%	53%	21%	.48 (.51)
Negative Affect	146	2.01 (0.66)	2.11 (0.81)	(+) 0.118	[−0.207, 0.443]	18%	58%	24%	.39 (.44)
Loneliness	667	2.24 (0.85)	2.27 (0.93)	(+) 0.030	[−0.122, 0.181]	8%	82%	10%	.45 (.55)
Organizational Trust	667	4.16 (0.72)	3.81 (0.94)	(−) 0.348*	[0.195, 0.500]	27%	63%	10%	.28 (.33)
Work Engagement	667	4.02 (0.83)	3.67 (1.01)	(−) 0.317*	[0.165, 0.470]	21%	73%	6%	.30 (.36)
Life Meaning	667	4.34 (0.68)	4.18 (0.85)	(−) 0.205*	[0.052, 0.357]	14%	79%	7%	.46 (.54)

Note. * $p < .05$.

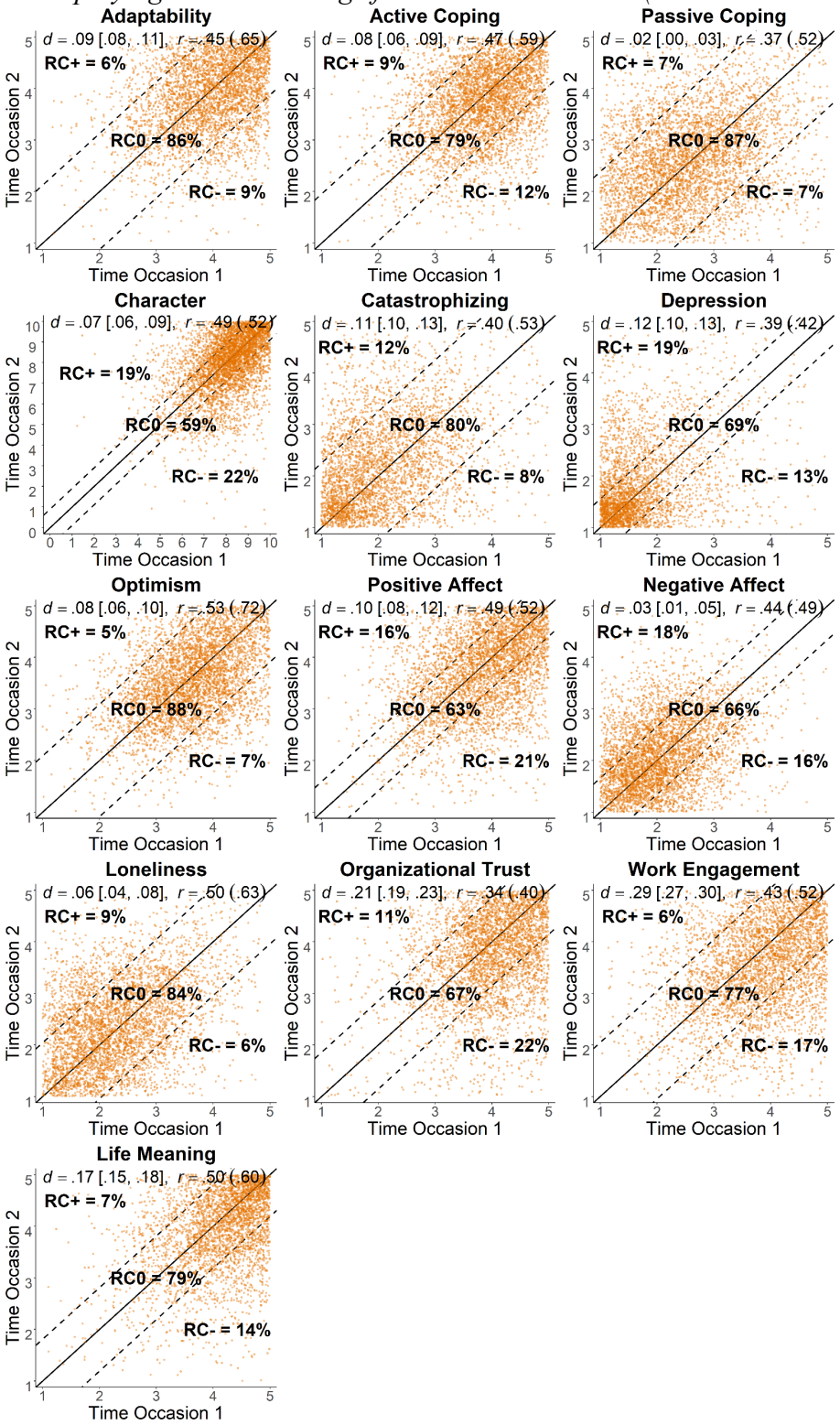
[a] Mean-level change in standard score units (i.e., Cohen's *d*). Absolute value is reported. Positive and negative signs in parentheses indicate direction of change.

[b] Percentage of individuals who decreased, had stable scores, or increased on each measure, according to reliable change (see text). The expected frequencies are 2.5%, 95%, and 2.5%, respectively.

[C] Correlations between Time Occasion 1 and Time Occasion 5. Values in parentheses are disattenuated for measurement error (based on Omega Total reliability).

Figure K3

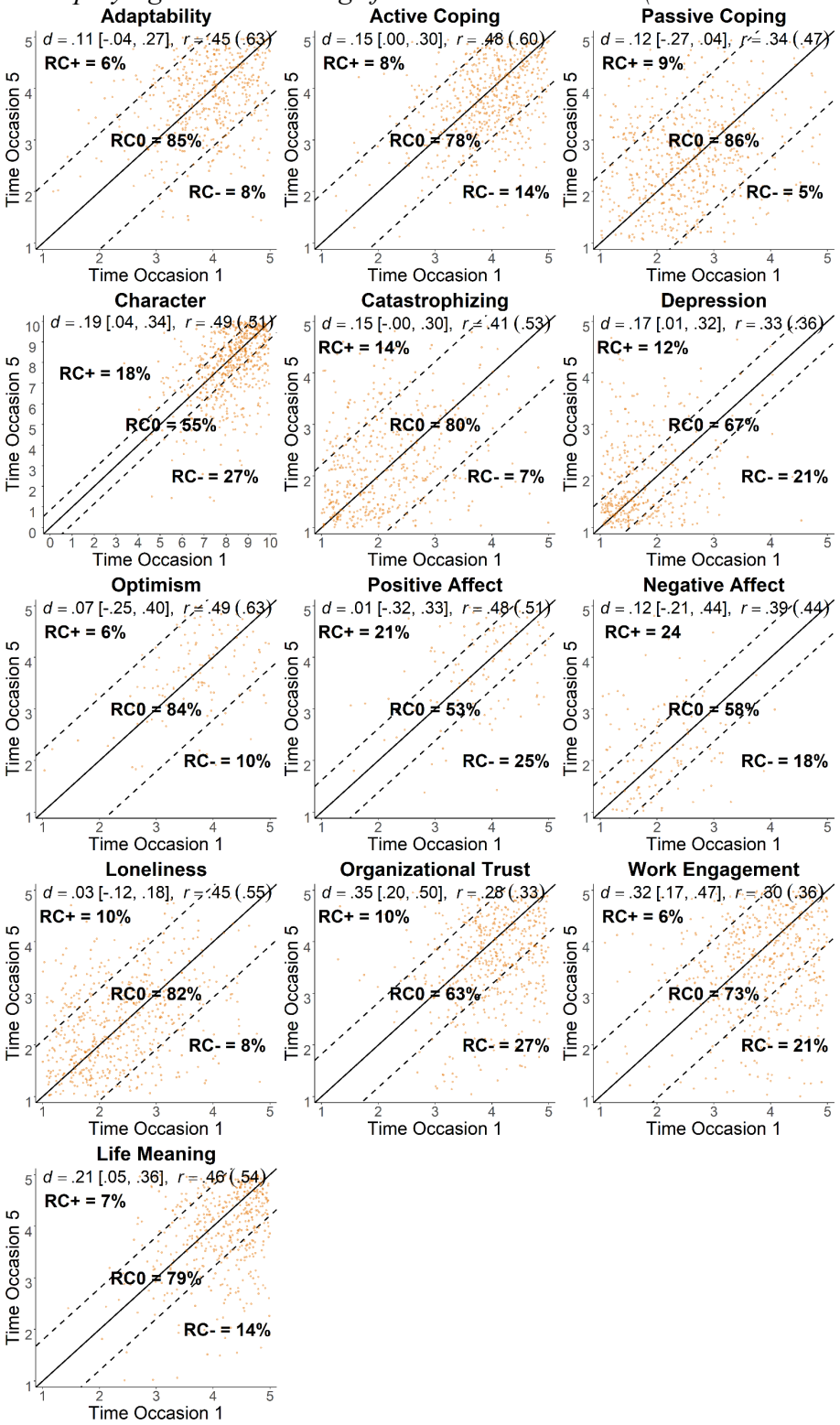
Modified Brinley Plots Displaying Reliable Change for GAT 2.0 Measures (Time 1 vs. Time 2)



Note. Modified Brinley plots showing individual-level change between Time Occasion 1 and Time Occasion 2. Points represent random subsample of 5,000 respondents (*n*s = 35,903–57,720). Diagonal solid lines = no change; diagonal dotted lines = reliable change (RC) upper and lower boundaries; RC+ = Reliable Change increase; RC− = Reliable Change decrease; RC0 = Indeterminate Change. *d* = Cohen's *d* with 95% confidence interval for mean-level change. *r* = rank-order stability correlation with attenuated correlation coefficient in parentheses.

Figure K4

Modified Brinley Plots Displaying Reliable Change for GAT 2.0 Measures (Time 1 vs. Time 5)



Note. Modified Brinley plots showing individual-level change between Time Occasion 1 and Time Occasion 5. Points represent sample of respondents (*n*s = 146–667). Diagonal solid lines = no change; diagonal dotted lines = reliable change (RC) upper and lower boundaries; RC+ = Reliable Change increase; RC− = Reliable Change decrease; RC0 = Indeterminate Change. *d* = Cohen's *d* with 95% confidence interval for mean-level change. *r* = rank-order stability correlation with attenuated correlation coefficient in parentheses.

Appendix L: Data Transparency and R Code

The data reported in this report were obtained with permission within the Person-Event Data Environment (PDE) managed by the Research Facilitation Laboratory (RFL) of the Army Analytics Group (AAG). The raw data and full scripts are available with proper research approvals through the PDE and RFL. First time users can access the PDE via the PDE portal (https://pde.army.mil/) using a Common Access Card. Once access has been granted to the PDE, interested parties may contact the corresponding author about how to obtain copies of the raw data and full R coding scripts contained within the PDE (a sample of scripts can be found in the supplemental materials). Researchers may also contact the corresponding author for further information about accessing the PDE (see also, Knapp et al., 2018).

R syntax code for the linkage of data used in this study can be found on our division's GitHub page (uva-bi-sdad) at the link below. Any questions about the data linkage or usage of the PDE can be directed to the corresponding author.

https://github.com/uva-bi-sdad/gat-stability