Measuring Psychological (In)flexibility: A Psychometric Evaluation and Comparison of Existing Measures

by

Kelsey N. Thomas

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
August 6, 2023

Keywords: psychological (in)flexibility, acceptance and commitment therapy, reliability, validity, contrast analysis

Approved by

Joseph Bardeen, Chair, Associate Professor of Psychological Sciences
Chris Correia, Professor of Psychological Sciences
Richard Macatee, Assistant Professor of Psychological Sciences
Brian Albanese, Assistant Professor of Psychological Sciences

Abstract

Psychological (in)flexibility, the tendency for behavior to be guided by psychological reactions rather than chosen values, is a central construct of interest in acceptance and commitment therapy (ACT; Hayes, Strosahl, & Wilson, 2012; Levin et al., 2014). Psychological (in)flexibility has received significant attention as a potentially important mechanism for understanding the development and maintenance of psychopathology. Although several self-report measures of this construct have been developed, some of the most widely used measures have been identified as having potentially serious problems. Additionally, there is a lack of consensus regarding which of these measures is most theoretically and psychometrically sound. In fact, the majority of recently developed measures of psychological (in)flexibility have not been independently evaluated. The overarching aim of the current study was to examine the reliability and validity of each of the identified measures of psychological (in)flexibility and provide further information about sensitivity of the measures to change over time. A sample of 474 adults were included in the sample at Time 1 and 127 participants were retained for a second assessment at Time 2. In general, results of reliability analyses suggested adequate internal consistency for most of the measures of psychological (in)flexibility that were examined in this study, though item redundancy emerged as an area of concern for several measures. Results of contrast analysis provided evidence that the Acceptance and Action Questionnaire-II (AAQ-II), Acceptance and Action Questionnaire-3 (AAQ-3), and Multidimensional Psychological Flexibility Inventory (MPFI) Inflexibility scale demonstrated the best fit within the specified nomological network. Despite these measures providing the best fit within the nomological network, their discriminant validity was questionable. Finally, results of sensitivity to change analysis indicated that most of the target measures were not particularly sensitive to changes in

psychological (in)flexibility over time. Findings from this study represent a step forward in refining the measurement of psychological (in)flexibility, and, in turn, improving understanding of the construct more generally.

Acknowledgments

I would like to express my deepest appreciation to my mentor, Dr. Joe Bardeen, for his mentorship throughout this project and the years leading up to this milestone. His patience, consistency, and dedication have been unparalleled, and I am grateful for his guidance. I would also like to thank my committee members, Dr. Chris Correia, Dr. Richard Macatee, and Dr. Brian Albanese, for their service and helpful guidance during the early stages of this project. Finally, this project would not have been possible without the instruction offered by Dr. Frank Weathers and Dr. Jess Petri, whose teaching enabled me to build the skills necessary to complete this study.

In addition to the Auburn faculty, I would like to acknowledge the role my family has played in shaping my dedication and drive to achieve at the highest academic levels. I am thankful for my husband and the day-to-day efforts of encouragement and reassurance that allowed me to progress through each stage of the dissertation process. And, I am grateful to my fellow students, with whom I have spent so much time over the last few years. Sharing this experience with them and benefitting from their companionship and support has been a constant source of motivation.

Table of Contents

List of Tables

Measuring Psychological (In)flexibility: A Psychometric Evaluation and Comparison of Existing Measures

Psychological flexibility has become a major focus of research across various forms of psychopathology and other relevant outcomes. The construct emerged from the acceptance and commitment therapy (ACT) literature, a transdiagnostic third-wave behavioral approach to psychological treatment. The foundation of ACT is in the framework of functional contextualism, a philosophical perspective that seeks to predict and understand behavior using empirically derived rules (Hayes, Pistorello, & Levin, 2012). From the perspective of functional contextualism, psychological flexibility is "the process of contacting the present moment fully and as a conscious human being and persisting or changing behavior in the service of chosen values" (Hayes et al., 2006, p. 9). Since its introduction, interest in the ACT model and psychological flexibility has increased steadily, with many researchers using the construct outside of the context of the ACT framework. A subset of the research in this area has also focused on accurate measurement of this construct. Given the popularity of the model in both applied and basic research, reliable and valid measurement of psychological flexibility is important.

The ACT Model

ACT approaches psychopathology from a largely behavioral perspective, with its guiding theory being grounded in functional contextualism, an approach to understanding behavior that emphasizes processes (e.g., the function of a thought) over form (e.g., thought content; Biglan & Hayes, 1996; Hayes, 1993). As the name suggests, a functional contextual approach to understanding behavior also accounts for the context in which the behavior occurs and considers the function of behaviors to be context dependent. More directly, functional contextualism

defines behavior as “the action of a whole organism within a particular context” (Villatte, Villatte, & Hayes, 2016). Further, the workability of behaviors is judged based on how well a behavior functions in a given context rather than labeling behaviors as strictly adaptive or maladaptive (Hayes, Pistorello, et al., 2012). Thus, understanding the function of behavior across contexts is vital to intervention within the ACT framework. Another important component of functional contextualism is the truth criterion, or the standard on which functional contextual science judges itself, with the goal being predicting and influencing effective action (Hayes et al., 1993). Thus, the truth criterion connects the philosophical perspective of functional contextualism to applied clinical intervention by providing the primary goal of responding to environmental cues that influence behavior in order to increase workable behavioral choices.

In addition to its unique approach to understanding behavior, ACT attempts to understand and address the basis of the types of psychological suffering that appear to be exclusive to humans. To this aim, ACT draws from Relational Frame Theory (RFT), which proposes that language is a learned behavior that is developed within and governed by socially based rules (Hayes et al., 2001). The ability to learn and use language allows humans to understand complex, symbolic relationships between stimuli. For example, humans are able to develop and apply relationships that are not based on the intrinsic or concrete characteristics of a given stimulus, such as associating a person with a specific name or identifying certain emotions as good or bad. In this way, language is thought to be a learned behavior that can influence future learning and behaviors (Villatte, Villatte, & Hayes, 2016). Over time, immensely complex relations develop between stimuli in the relational frame network. As a result, the function of behavior is typically predicated by this complex network of linguistic relations elicited by contextual cues in the environment rather than direct environmental drivers (Hayes, 1989). Taken together, with the

emphasis on function and context, the linguistic relationships proposed within RFT provide an alternative target of intervention in contrast to existing behavioral approaches. The ineffective application of these relational frames within a given context (e.g., increased heartrate after climbing stairs being associated with catastrophic heart failure, a common association among those with panic disorder) is the target of intervention. That is, the way that individuals interact with internal experiences (e.g., thoughts, emotions) is the target of intervention rather than attempting to directly change those internal experiences (e.g., challenging thought content; Twohig, 2012). If individuals are able to change the way they utilize language and cognition to interact with the world, they may also be able to engage in more workable, effective behaviors.

In order to translate the theoretical basis of ACT into an actionable treatment, a model has been developed that focuses on psychological flexibility. Psychological flexibility is a multidimensional construct that is comprised of six core processes: acceptance, defusion, present moment awareness, self-as-context, values, and committed action. This multidimensional framework is most often referred to as the ACT Hexaflex (Hayes, Pistorello, et al., 2012). Theoretically, each of these components are highly related but distinct constructs. The Hexaflex can also be broken down into three dyads representing opening up to experiences (acceptance and defusion), being present, (present moment awareness, self-as-context), and acting in a value-consistent manner (values, committed action). Additionally, some conceptualizations of psychological flexibility include an opposite inflexibility construct that consists of six components directly corresponding to the Hexaflex components: experiential avoidance, fusion, lack of contact with the present moment, self-as-content, lack of contact with values, and inaction.

Debate exists as to the nature of the relationship between psychological flexibility and inflexibility, with some researchers considering the two constructs to be polar opposites of a single spectrum, while others view the two as related but independent (Bond et al., 2011; Cherry et al., 2021; Rolffs et al., 2018). Evaluation of this question has been somewhat limited given that, until recently, separate scales of flexibility and inflexibility have not existed. However, recent research has clearly demonstrated distinction between the two constructs (Kashdan & Rottenberg, 2010; Rolffs, et al., 2018; Stabbe et al., 2019). Therefore, when terms are used broadly or non-specifically, these two constructs will be represented with the term (in)flexibility throughout this proposal.

Practically, psychological flexibility represents skills that are essential for living with the suffering that is inevitable in life, while psychological inflexibility might be thought of as the absence of those skills. Within RFT, psychological inflexibility is thought to be the outcome of the interaction between problematic cognition and rigid language that limits behavior such that individuals are unable to engage in value-consistent behavior in the long-term (Hayes et al., 2006). As such, the ACT model of psychotherapy attempts to reduce inflexibility by changing the way individuals use language and interact with their own thoughts in order to increase values-directed behavior. Within the context of treatment, aspects of the Hexaflex are often applied together and build on each other over time. However, the distinct processes have also been evaluated independently to varying degrees. Many researchers have called for dismantling studies in order to better understand the unique contribution of each construct, but these types of studies are difficult to carry out and few have examined the entirety of the ACT model. Despite this, the six theoretical processes underlying the ACT Hexaflex have been described and researchers have examined how these constructs may relate to various forms of psychopathology

(e.g., Bramwell & Richardson, 2018; Forman et al., 2012; Levin, Krafft, & Twohig, 2020; Stockton et al., 2019).

One of the most prominent aspects of the ACT model and psychological (in)flexibility, and the construct that has received the most empirical attention, is experiential avoidance. Experiential avoidance is described as an unwillingness to come into contact with potentially aversive or unwanted internal experiences (Hayes et al., 1996). Experiential avoidance has been connected to a number of specific and general (e.g., negative affect) maladaptive psychological outcomes (Chawla & Ostafin, 2007; Levin et al., 2018; Wenze et al., 2018). In contrast to experiential avoidance, acceptance represents being open to all internal experiences (Hayes et al., 2004). Being accepting also involves choosing not to struggle with and attempt to change such experiences. While acceptance is a major component of the ACT model, it is important to note that acceptance is a facilitator of values-directed behavior, rather than being the goal itself. Fewer studies have examined the relationship of acceptance with relevant outcomes. Current research suggests that acceptance is more effective in regulating emotions than strategies associated with non-acceptance (e.g., suppression; Hofmann et al., 2009). Further, researchers have suggested that acceptance and experiential avoidance may mediate ACT treatment outcomes, and thus, are essential ingredients for alleviating emotional suffering (Forman et al., 2007).

Cognitive defusion is another component of being open and willing to potentially aversive experiences. Cognitive defusion is the process of disentangling oneself from one's own thoughts and interacting with those thoughts in a non-judgmental manner (Hayes, Pistorello, et al., 2012). ACT proponents have argued that taking thoughts literally, or being cognitively fused, is one important source of distress that may interrupt an individual's ability to engage in values-

directed action (Hayes, Pistorello, et al., 2012). From a theoretical perspective, cognitive fusion creates a situation where choices are made primarily based on the language assigned to internal events by individuals, while other important contextual information derived from direct experience is typically disregarded. While relatively less research has been devoted to understanding fusion and defusion compared to experiential avoidance, research has demonstrated that cognitive defusion exercises (e.g., repeating a word until it loses its meaning) perform as well or better than comparable cognitive strategies such as cognitive restructuring in lessening emotional distress (Deacon et al., 2011; Masuda et al., 2010; Pilecki & McKay, 2012). The negative impact of cognitive fusion has also been well-documented, with numerous studies demonstrating relationships between fusion and poorer mental health outcomes (e.g., Fergus, 2015; Krafft et al., 2019; Trindade & Ferreira, 2014).

In addition to openness, the ACT model and psychological (in)flexibility construct emphasize awareness. More specifically, the Hexaflex includes the present moment awareness and self-as-context constructs. Present moment awareness, or contact with the present moment, involves active, non-judgmental awareness of both psychological and environmental experiences (Hayes et al., 2006). Researchers have highlighted the tendency to rigidly focus on the future or past as problematic due to the inability to engage in values-directed behavior when the mind is occupied with correcting mistakes or planning for future problems. Further, efforts to avoid contact with the present moment such as rumination have been tied to negative outcomes such as depressed mood (Davis & Nolen-Hoeksema, 2000). Contact with the present moment has also been conceptualized as a form of exposure, in that individuals are purposefully coming into contact with feared stimuli (Hayes, 2004). This conceptualization is consistent with both the ACT framework as well as behavioral approaches to treating psychological disorders in general.

The contact with the present moment construct is consistent with the emphasis in other third-wave cognitive-behavioral models on mindfulness and awareness. Within the ACT-context, research has indicated that present moment awareness may be an important predictor of symptom change during treatment (McCracken & Gutiérrez-Martínez, 2011).

Self-as-context, or the process of becoming the contextualized self, has received relatively little attention in the extant literature compared to other Hexaflex components. Put simply, self-as-context is one way of conceptualizing self-experience that emphasizes the distinction between internal experiences (e.g., cognitions and emotions) and sense of self (Hayes et al., 2011). The opposite of self-as-context is represented by the conceptualized self, a perspective in which individuals are overly attached to the concept of self to the point that they prioritize the preservation of that concept and have difficulty taking different perspectives in order to engage in effective values-driven action (Twohig, 2012). Instead, by viewing oneself as the context, ACT theorizes that individuals can engage in non-judgmental awareness of their thoughts as well as other internal experiences (Hayes et al., 2006). The construct is thought to be closely related to mindfulness and is typically integrated into mindfulness exercises during treatment (Hayes et al., 2012). More recently, researchers have questioned if self-as-context can be effectively integrated into treatment, especially as a standalone process (Godbee & Kangas, 2020). Further, research in this area has been somewhat limited because measures of this construct were only recently developed.

The final dyad of processes in the ACT model focuses on doing what matters by engaging in values-directed behavior. Within the ACT context, values are freely chosen, self-selected reinforcers that facilitate moment-to-moment behaviors and longer-term goals (Wilson et al., 2010). Importantly, values are considered dynamic and are explored and evaluated in an

ongoing process throughout ACT. Thus, in contrast to temporary, attainable goals, values provide the motivation and intrinsic drive to engage in action over the long-term (Serowik et al., 2018). Individuals who lack contact with their values are thought to also lack direction or clear motivation for their behaviors, which may lead to reactive choices and a less fulfilling life overall (Hayes et al., 2006). From a theoretical perspective, values alter the reinforcement value of certain behavioral choices, leading to increased values-directed behavior, and, potentially, increased psychological flexibility (Twohig, 2012). Although values are central to psychological flexibility, few studies have evaluated their unique contribution to treatment outcomes. Initial evidence of studies that have isolated the effects of values clarification on treatment outcomes are promising, with results indicating that changes in values precede reductions in distress (Gloster et al., 2017; Levin et al., 2020).

While values are largely internally held, committed action represents the translation of those values into active behavioral choices. Efforts to engage in committed action involve utilizing each of the other five psychological flexibility processes in order to meet progressively larger and more complex goals (Hayes et al., 2006). In the context of treatment, many types of interventions (e.g., exposure, skills training, behavioral activation) can be utilized to increase committed action if they are performed in a way that is consistent with the ACT model (Twohig, 2012). These types of therapeutic efforts allow individuals to come into contact with their own discomfort in a manner that can be addressed with the other ACT processes. The unique contribution of changes in committed action to changes in overall psychological flexibility has not frequently been studied relative to the other Hexaflex components. The little evidence that exists suggests that committed action plays an important role, in combination with several other Hexaflex processes, in producing positive changes in global flexibility (Levin, Krafft, & Twohig,

2020; Levin et al., 2020; Villatte et al., 2016). Further, values-based action seems to play a role in improving emotion and physical functioning, though these effects have primarily been studied in chronic pain samples (Vowles et al., 2011).

While recent research has begun to consider the unique contributions of each of the six Hexaflex processes to psychological flexibility and other psychological processes, much of the research on psychological (in)flexibility has focused on the global construct(s) rather than the specific lower-order domains outlined in the Hexaflex model. With regard to clinical applications, several meta-analyses have provided support for ACT across a range of presenting concerns (e.g., Arch et al., 2012; Bai et al., 2020; Benfer et al., 2021; Gloster et al., 2020). Others have suggested that current evidence is still in initial stages and is not strong enough to classify ACT as being even moderately efficacious (Hacker et al., 2016; Öst, 2014). The most prominent finding across the meta-analyses that have been conducted is that ACT tends to result in greater treatment response when compared to control conditions, but not when compared to active treatments such as cognitive-behavioral therapy (CBT), though some research has suggested that ACT and CBT may be equivalent with regard to treatment outcomes (A-Tjak et al., 2015; Dimidjian et al., 2016; Gloster et al., 2020; Twohig & Levin, 2017). Taken together, it seems that there is preliminary support for using ACT for a wide variety of psychological problems.

Given that ACT is a transdiagnostic treatment, a wide variety of measures have been developed to capture changes in ACT-related processes. The Acceptance and Action Questionnaire-II (AAQ-II; Bond et al., 2011) is the most widely used measure of global psychological (in)flexibility (Gloster et al., 2020). Despite widespread use of the AAQ-II, questions about the psychometric properties of the measure, as well as interest in alternative

methods of capturing psychological (in)flexibility, have driven an exponentially growing body of literature on measurement of psychological (in)flexibility from the ACT perspective.

Assessment of ACT constructs

Recent interest in measurement of psychological (in)flexibility has resulted in a rapidly expanding literature with numerous options for capturing global psychological (in)flexibility as well as the underlying processes described above. As was noted above, the AAQ-II has been the most widely utilized measure of psychological (in)flexibility. The AAQ-II was developed with the goal of improving the psychometric properties of the original AAQ (Bond et al., 2011; Hayes et al., 2004). While the AAQ represented an important first step in furthering research related to psychological (in)flexibility, the measure exhibited questionable psychometric properties, including low to poor internal consistency and low test-retest reliability over a period of several months (Hayes et al., 2004; Zvolensky et al., 2005). Additionally, the AAQ lacked construct clarity. Both the measure itself and conceptual information offered by the authors of the measure lead to confusion regarding the nature of experiential avoidance, which was what the measure was originally designed to assess. More specifically, researchers questioned if the construct should be considered unidimensional or multidimensional and had difficulty distinguishing between experiential avoidance and similar constructs such as suppression or general psychological distress (Chawla & Ostafin, 2007; Zvolensky et al., 2005). Finally, the factor structure of the AAQ was thought to be unstable as evidenced by several studies finding multi-factor structures as opposed to the original unidimensional factor structure (Bond & Bunce, 2003; Bond et al., 2011). Thus, the AAQ-II was developed in an attempt to refine measurement in the ACT literature.

In general, the AAQ-II is considered a substantial improvement compared to the AAQ. It is more reliable with high internal consistency and modest test-retest reliability over a period of several months (Bond et al., 2011; Fledderus et al., 2012). Additionally, the AAQ-II exhibits criterion-related validity with relevant clinical constructs such as depression and anxiety (Pennato et al., 2013). The measure also consistently demonstrates a unidimensional factor structure across diverse samples, though this remains in contrast to the multidimensional model of psychological (in)flexibility proposed by ACT (Fledderus et al., 2012; Pennato et al., 2013; Ruiz et al., 2016). While the AAQ-II is a significant improvement over the AAQ, several criticisms remain unresolved.

First and most prominently, the AAQ-II tends to fail to distinguish between psychological (in)flexibility and related but distinct constructs such as general emotional distress and negative affect (Rochefort et al., 2018; Vaughan-Johnston et al., 2017; Wolgast, 2014). While a relationship between psychological (in)flexibility and constructs of emotional distress would be expected, the large size of the magnitude of these relations have raised concerns that the AAQ-II may not be measuring a distinct construct. The idea that the majority of variance in AAQ-II outcomes is attributable to negative affect may be quite problematic because of the substantial role played by negative affect in most forms of psychopathology (Barlow et al., 2004). Additionally, with regard to criterion-related validity, there is an apparent lack of specificity regarding what psychological (in)flexibility will predict (Gámez et al., 2011; Rochefort et al., 2018; Wolgast, 2014). That is, rather than a clear set of criterion measures that would help establish construct validity, the AAQ-II exhibits large magnitude correlations with a wide variety of constructs, thus resulting in a convoluted nomological network.

Another prominent criticism of the AAQ-II is that it lacks construct clarity. In the process of revising the AAQ, the authors of the AAQ-II updated their conceptualization such that the measure is said to capture psychological (in)flexibility or experiential avoidance. This step may have reflected an attempt to clarify the potential uses of the measure; however, the conflict with theory and the authors interchangeable use of the terms psychological (in)flexibility and experiential avoidance has led to continued confusion in the literature (Rochefort et al., 2018). As a result, the AAQ-II has been criticized for contributing to the lack of clarity in the definition of psychological (in)flexibility (Francis et al., 2016). Evidence has indicated that the AAQ-II does not converge with alternative measures of experiential avoidance, but due to the synonymous use of psychological (in)flexibility and experiential avoidance, some researchers continue to use the AAQ-II as a measure of the latter (Tyndall et al., 2019). Additionally, while the factor structure of the AAQ-II is thought to be quite stable, the single composite score representing the level of psychological (in)flexibility does not provide information about the proposed underlying components of the construct from the ACT Hexaflex model.

In response to the recent widespread criticism of the AAQ-II, Ong, Pierce, Woods, Twohig, and Levin (2019) utilized item response theory in an attempt to address some of the psychometric concerns described above. Results of this study demonstrated that AAQ-II items varied in their ability to differentiate among individuals with lower levels of psychological (in)flexibility. Additionally, most of the items had decreased utility at low levels of (in)flexibility. The authors suggested that the use of broad, nonspecific language (e.g., emotions or feelings) effects the relative accuracy of the items with regard to assessing psychological (in)flexibility. Based on the results of this study, Ong et al. (2020) developed the Acceptance and Action Questionnaire-3 (AAQ-3), a revised version of the AAQ-II that consists of the same

items with modified wording to increase clarity and potentially increase sensitivity and discrimination of the items as well as discriminant validity with regard to measures of general distress. The AAQ-3 exhibited slightly improved discriminant validity compared to the AAQ-II in community and student samples, though neither measure effectively discriminated between psychological (in)flexibility and general distress in a treatment seeking sample (Ong et al., 2020). With regard to item performance, the AAQ-II and AAQ-3 performed similarly, with both measures demonstrating relatively higher sensitivity to individual differences in a community sample rather than student or clinical samples. While these results do not suggest that the revision is a substantial improvement, the AAQ-3 is likely worth further examination, especially given the popularity and ease-of-use associated with its predecessor.

In addition to revising the AAQ-II, other authors have developed alternative measures of psychological (in)flexibility that differ more substantially from their predecessor. Newer measures of psychological (in)flexibility have taken several different approaches to representing the theoretical model, but generally characterize psychological (in)flexibility as multidimensional. This is in sharp contrast to the unidimensional nature of the AAQ-II, though the ACT theoretical model does fit well with a multidimensional approach. Additionally, multidimensional approaches are often seen as more practical and efficient as test users can collect a large amount of information with one tool rather than many. Despite these potential benefits, some researchers have questioned the utility of these types of multidimensional measures, especially if there is a lack of theoretical clarity regarding how the lower-order constructs should relate to one another and to relevant criterion (Johnson et al., 2012). The most prominent multidimensional self-report measures for assessing psychological (in)flexibility are the Multidimensional Psychological Flexibility Inventory (MPFI) and the Comprehensive

Assessment of Acceptance and Commitment Therapy Processes (CompACT); both of these measures have demonstrated preliminary evidence of adequate construct validity relative to the AAQ-II (Francis et al., 2016; Rolffs, et al., 2018). While this preliminary evidence is promising, it is important to note these measures vary slightly in their conceptual models of psychological (in)flexibility, and as a result, the number of domain-specific factors differ between the measures.

The Personalized Psychological Flexibility Inventory (PPFI) seeks to account for individual differences in psychological (in)flexibility with specific regard to individually dictated values-based action (Kashdan et al., 2020). The authors of the PPFI argue that the inclusion of values identification allows the tool to align more closely with the ACT model than previous measures. This is a somewhat unusual approach to measurement as idiographic techniques can be difficult to integrate into nomothetic models (Wright & Zimmerman, 2019). Further, idiographic approaches have traditionally been reserved for single-case designs or methods that collect large amounts of data from each participant such as ecological momentary assessment. Researchers have recently highlighted the potential contribution of the integration of idiographic elements with more traditional assessment approaches, suggesting that it is possible to develop reliable and valid tools that are able to capture individual differences more adequately (Wright & Woods, 2020). With specific regard to the PPFI, the authors suggest that this combined approach may alleviate some of the challenges that previous measures have faced such as a lack of sensitivity to individual differences (Kashdan et al., 2020).

While researchers have utilized a number of different approaches to measuring psychological (in)flexibility, efforts have been primarily focused on improving discriminant validity over what is offered by the AAQ-II. In contrast, few of these measures have prioritized

improving reliability in the measurement of psychological (in)flexibility and its components. Additionally, most of these measures have not been independently evaluated beyond their initial development but are receiving widespread recommendations for continued use without additional psychometric testing. For example, the MPFI was used to develop latent profiles of psychological (in)flexibility in order to provide nuanced treatment recommendations to clinicians (Stabbe et al., 2019). Further, several researchers have recommended the use of multidimensional measures of psychological (in)flexibility while making specific reference to the MPFI and CompACT (e.g., Levin et al., 2020; Luoma et al., 2019; Ong et al., 2020; Stockton et al., 2019). The use of these measures without first conducting adequate psychometric evaluation could lead to potentially misleading results. Clark and Watson (2019) characterized scale development and validation as an iterative process that only truly ends when a scale ceases to be used, illustrating the value in collecting substantial evidence of validity in order to feel confident in a given scale. Further, it is important to ensure that the development and evaluation of a given scale conforms to the assumptions of Classical Test Theory as much as possible in order to ensure that measurement error is minimized. For example, ensuring that a test is reliable across time and across samples is a basic element of Classical Test Theory and a necessary condition for establishing validity. Researchers often rely on internal consistency, most often Cronbach's alpha, as a measure of reliability but fail to recognize the limitations of such an approach. Specifically, Cronbach's alpha may be problematic for some scales as it assumes unidimensionality and is vulnerable to inflation when a scale has many items (Hayes & Coutts, 2020). Within the psychological (in)flexibility literature, researchers have occasionally considered test-retest reliability as well. Specifically, the authors of both the AAQ-II and PPFI examined test-retest reliability using correlation coefficients between the measures given several

months apart (Bond et al., 2011; Kashdan et al., 2020). These estimates suggest acceptable to good retest reliability (.60 to .80), but Pearson correlations are not ideal given their tendency to vary based on the time between measurements and vulnerability to sample variance (Koo & Li, 2017). As described, reliability is a vital aspect of test construction within the Classical Test Theory framework. Evidence of reliability provides the basis on which the validity of a test can be more clearly established.

Sensitivity to change, or the ability to accurately capture change over time, is an important aspect of test construction and evaluation that is often overlooked, especially within the context of clinical applications (Fok & Henry, 2015; Vermeersch et al., 2000). Sensitivity to change is often considered to be a form of elaborative validity (Foster & Cone, 1995; Hays & Hadorn, 1992; Stratford, Binkley, & Riddle, 1996). Specifically, if theory suggests that a specific construct is fluid under certain circumstances, or simply changes as a function of time, it is important to test whether scores on a specific measure fluctuate as would be expected. The usefulness of a given measure may depend on that measure's ability to adequately detect significant and clinically meaningful differences. Measures that have not been evaluated for responsiveness may have limited utility in research or clinical applications (Stratford, Binkley, & Riddle, 1996). Scores on measures of psychological (in)flexibility are used in clinical trials and to track treatment progress for individual clients. As such, it will be important to examine sensitivity to change of widely used measures of psychological (in)flexibility.

At this time, only two studies have addressed this concern with regard to measures of psychological (in)flexibility. Rolffs et al. (2018) evaluated the sensitivity to change of the MPFI when it was originally developed, with the goal of determining how much an individual's score would need to change on each scale to detect a significant change given the measurement error

of that scale. According to Rolffs et al., an MPFI score would only need to change by approximately one to two points in order to be able to capture significant change. Rolffs et al. (2018) noted that the MPFI is able to detect change slightly more effectively than the AAQ-II while providing substantially more information. A second investigation of this measurement characteristic was conducted by Benoy et al. (2019) and included a comparison of the AAQ-II as well as the Open and Engaged State Questionnaire and the Psyflex, two less widely used measures of psychological (in)flexibility (Benoy et al., 2018; Firsching et al., 2018). In contrast to the approach used by Rolffs et al. (2018), Benoy et al. (2019) considered sensitivity to change following an intervention. When compared to alternative measures of psychological (in)flexibility, the AAQ-II underperforms with regard to sensitivity to change following an intervention (Benoy et al., 2019). Based on this research, it seems possible that the relatively lower sensitivity of the AAQ-II may result in underestimated or under-identified effects as it has demonstrated poorer sensitivity to change following an intervention and high minimally detectable changes over time. Additional research examining sensitivity to change across a wider variety of commonly used measures of psychological (in)flexibility is necessary to clarify and extend existing findings.

As described, sensitivity to change has received relatively little attention in the development of measures of psychological (in)flexibility. In fact, most measures of psychological (in)flexibility are developed with cross-sectional data. The assumption that a scale developed solely with cross-sectional data will detect changes in the construct of interest over time is flawed (Husted et al., 2000). With regard to short-term laboratory studies, a lack of sensitivity limits the utility of most measures of psychological (in)flexibility and could contribute to difficulty identifying meaningful effects in treatment outcome research. Further, sensitivity to

change for the AAQ-II and other measures of psychological (in)flexibility may be negatively impacted by the use of broad, nonspecific language. Because such items can be interpreted in a wide variety of ways, there may be a general lack of consistency in responding across participants who have similar levels of the construct of interest. For example, Benoy et al. (2019) noted that a slight change in AAQ-II wording (i.e., "I am afraid of my feelings" to "I am afraid of my feelings such as anxiety, panic, depression, etc.") resulted in substantially greater increases in psychological flexibility over the course of treatment. These types of content changes are frequently included in efforts to revise existing measures (see Ong et al., 2019). Additionally, lack of clarity in item content has been frequently cited as problematic, such that it is difficult to determine if responses reflect beliefs about thoughts and emotions or if they reflect the degree to which these internal events are aversive (Conner & Barrett, 2012; Gámez et al., 2011). Other researchers have noted the importance of context in accurately capturing psychological (in)flexibility, which may also impact the ability to collect accurate measurements across time (Hayes, Barnes-Holmes, & Wilson, 2012; Kashdan & Rottenberg, 2010).

In sum, research on measurement of psychological (in)flexibility and related constructs has increased exponentially in recent years. Given the multitude of new measures and general lack of independent evaluation, the current study will provide an evaluation of the psychometric properties of several emerging measures of psychological (in)flexibility. The primary aim of the present study was to examine the construct validity of these measures with particular emphasis on discriminant validity. To address this aim, the relations between scores on each measure and criterion variables were examined to determine whether each measure is situated properly within the nomological network as would be expected based on theory (i.e., examining the convergent-discriminant array). The degree to which a score accurately represents the construct of interest

can be assessed by comparing theoretically expected correlations between scores on the target measure (i.e., the construct of interest) and scores on a wide variety of criterion variables to empirically derived correlations between the target measure and criterion variables (i.e., contrast analysis; Westen and Rosenthal; 2003). This procedure is intended to reduce imprecise interpretations of correlations in favor of a more systematic approach (Furr & Heuckeroth, 2019). Criterion variables included measures of neuroticism, negative affect, social desirability, empathy, and various forms of psychopathology.

With regard to validity, predictions were based on both a theoretical basis and based on existing literature. Given that each of the measures of psychological (in)flexibility are purported to capture the same construct, predicted correlations with each of the criterion measures were predicted to be similar. One exception was the MPFI Flexibility composite, which was likely to differ somewhat given that evidence suggests that this scale is distinct from inflexibility rather than simply being its polar opposite (Rolffs et al., 2018). I expected that correlations for this scale with criterion measures would reflect the relatively weaker correlations than measures of this construct have exhibited with measures of psychopathology compared to correlations between psychological inflexibility and measures of psychopathology (Stabbe et al., 2019). Additionally, it was hypothesized that the more recently developed measures of psychological (in)flexibility (i.e., MPFI, CompACT, and PPFI), which were developed with the specific goal of improving content validity relative to the AAQ-II, would likely exhibit small magnitude associations with empathy and social desirability and medium to large magnitude associations with measures of psychopathology (Bond et al., 2011; Francis et al., 2016; Kashdan et al., 2020; Thompson et al., 2019). Additionally, it was expected that these measures would outperform the AAQ-II with regard to demonstrating correlations of a smaller magnitude with measures of

neuroticism and negative affect; that is, they would exhibit better discriminant validity. In other words, it was expected that these measures would align more closely with theoretically predicted correlations, as evidenced by $r_{alerting\text{-}CV}$ and $r_{contrast\text{-}C} > .70$, (Westen & Rosenthal, 2003). It was also predicted that the AAQ-3 would exhibit modest improvements in validity compared to the AAQ-II based on preliminary evidence which suggests that the AAQ-3 slightly outperformed the AAQ-II with criterion measures of depression, anxiety, and stress in some samples (Ong et al., 2020).

Reliability was examined prior to conducting contrast analysis because adequate reliability is a necessary condition for establishing adequate validity. In order to accommodate the variety of approaches utilized in constructing newer measures of psychological (in)flexibility, several robust indicators (i.e., McDonald's ω and the Intraclass Correlation Coefficient) were utilized to examine the internal consistency of the identified measures. Test-retest reliability of each measure was also examined. Based on existing evidence, it was hypothesized that reliability would be better for measures with greater, rather than less, content coverage (e.g., MPFI compared to the AAQ-II). However, previous research suggests that all of the measures that were examined in this study would exhibit acceptable reliability (Bond et al., 2011; Francis et al., 2016; Kashdan et al., 2020; Rolffs et al., 2018).

An additional aim of this study was to evaluate the potential for measures of psychological (in)flexibility to detect change across time when accounting for measurement error. This is especially pertinent given the purported role of psychological (in)flexibility in ACT-related outcomes. Given that the ability of most measures of psychological (in)flexibility to detect changes over time is relatively unknown, evaluating this aspect of recently developed measures may provide researchers and treatment providers with guidance with regard to measure

selection (Husted et al., 2000). While psychological (in)flexibility is generally thought to be relatively stable over time, several researchers have promoted the emerging idea that the construct may be more dynamic and state- or context-dependent than previously thought (Cherry et al., 2021; Doorley et al., 2020; Rolffs et al., 2018). At this time, it remains unclear how much change would need to occur to be able to detect significant and clinically meaningful change or progress on most of the measures being evaluated here. It was hypothesized that measures with greater content coverage that are more theoretically consistent with the ACT definition of psychological (in)flexibility (i.e., MPFI, CompACT, and PPFI) would exhibit greater sensitivity to change relative to the AAQ-II and AAQ-3.

Methods

Participants and Recruitment

Participants were adults from the United States between the ages of 18 and 64 and were recruited from Amazon's Mechanical Turk (MTurk), a crowdsourcing marketplace that allows researchers to directly connect with participants who engage in studies in exchange for compensation. Previous studies have consistently provided evidence of high-quality data collected via MTurk samples (Behrend et al., 2011; Buhrmester et al., 2011; Chandler & Shapiro, 2016; Hauser & Schwarz, 2016; Paolacci et al., 2010; Shapiro et al., 2013). Further, evidence suggests that when data quality control measures are taken, participants in research conducted using MTurk are more attentive and more diverse compared to typical college student samples (Behrend et al., 2011; Chandler & Shapiro, 2016; Hauser & Schwarz, 2016). These types of sample characteristics lend themselves to the study of psychological constructs such as psychological (in)flexibility as they provide a higher degree of certainty that the sample will adequately represent all levels of the variables of interest.

Several data collection strategies were implemented based on existing guidance for ensuring high quality MTurk data collection. First, MTurk workers were required to have an approval rating of at least 95% for completion of at least 50 human intelligence tasks. These ratings are provided by researchers based on the successful and reliable completion of a study task by participants. Restricting participation based on this criterion yields higher quality data than studies that rely on alternative quality control methods such as embedded catch questions (Peer et al., 2014). Further, authentication of responses as genuine was ensured by the inclusion of a Captcha verification question (Yarrish et al., 2020). Per recent recommendations made by Chmielewski and Kucker (2020), no additional strategies to ensure validity during the data collection were utilized in an effort to minimize potential burdens on participants. Instead, completion time and patterns of consecutive responding were examined in order to identify potentially careless responders (Curran, 2016; Wood et al., 2017).

Sample Size

When available, existing sample size recommendations for reliability and validity analyses tend to vary widely, with a range of 200 to 400 participants being most common (Clark & Watson, 2019; Guadagnoli & Velicer, 1988). Notably, studies using the Westen and Rosenthal (2003) approach to construct validity often report samples of approximately 400 to 500 participants (e.g., Carretero-Dios et al., 2011; Doi et al., 2018; Petri et al., 2020; Williams, et al., 2018). Given the variability in estimates and the relatively large number of relationships being examined in the present study, a minimum of 450 participants was deemed necessary to complete the first assessment session (Time 1). Special consideration was given to accounting for the possibility of attrition given the longitudinal nature of the proposed study. Attrition from MTurk studies can vary widely. However, in a recent review of attrition in internet-based

studies, the majority of studies reported drop-out rates of approximately 30% (Zhou & Fishbach, 2016). Study characteristics such as item difficulty and length of the self-report battery are important predictors of participant drop-out (Buhrmester et al., 2018; Crump et al., 2013). Individual characteristics such as motivation and education-level may also play a role in attrition (Zhou & Fishbach, 2016). As such, this study will be designed with the goal of limiting participant burden while promoting transparency and feasibility. With specific regard to longitudinal MTurk studies, the literature suggests an average completion rate of approximately 60% at initial follow-up, with diminishing rates of completion with greater lengths of time between study sessions (Daly & Nataraajan, 2015; Keith, Tay, & Harms, 2017). Daly and Nataraajan (2015) suggest that increased payment levels from the first to the second assessment session may improve response rates. In the proposed study, increased payment at the second assessment session was utilized to promote study completion. With an anticipated attrition rate of approximately 30-40%, I expected the Time 2 sample to consist of between 270 to 315 participants.

An initial sample of 504 participants completed the Time 1 survey on MTurk. Completion time and patterns of consecutive responding were examined in order to identify potentially careless responders as described above. Eight individuals were excluded based on a minimum time requirement of 618 seconds (n = 496), which was based on the two second per item cutoff recommended in the literature (Huang et al., 2012). Additionally, based on precedent in the literature that a high number of consecutive responses may suggest careless responding (Curran, 2016), long string analysis was used to exclude cases in which participants gave the same response for 30 consecutive survey items (n = 22). The final Time 1 sample (N = 474) was primarily male (55.3%), with an average age of 36.56 (SD = 9.55; *range* = 19-64). The majority

of the sample reported their race as White or European American (84.4%), followed by Black or African American (12.2%), Asian (1.5%), and American Indian or Alaskan Native (1.3%). A minority of the sample (20.7%) identified themselves as Hispanic or Latino/a.

Attempts were made to contact all participants from the Time 1 study, regardless of their inclusion in the final sample. Of the Time 1 participants, 291 (57.74%) were unable to be contacted due to no longer having an active MTurk account. One-hundred and thirty-six participants completed the Time 2 survey. The approach for identifying careless responders at Time 1 was used at Time 2. Five individuals were excluded from the Time 2 sample based on the minimum time requirement of 618 seconds (n = 131). An additional four individuals were excluded from further analyses due to having the same response for at least 30 consecutive items. The final Time 2 sample (N = 127) was primarily male (53.5%), with an average age of 37.24 (SD = 9.74; *range* = 21-64). The majority of the sample reported their race as White or European American (81.1%), followed by Black or African American (13.4%), Asian (3.1%), and American Indian or Alaskan Native (0.8%). A minority of the sample (18.1%) identified themselves as Hispanic or Latino/a.

Procedure

Study procedures were approved by the local Internal Review Board. Participants were able to complete the study from any computer with internet access. Interested participants were presented with a brief description of the study, including an estimation of the amount of time that it would take to complete the study and information about the longitudinal nature of the investigation. If they wished to proceed, participants provided electronic informed consent prior to beginning the study. By providing this consent, participants also consented to be contacted by the primary investigator to participate in the second phase of data collection. Individuals who

participated in the study received financial compensation ($1.50), which was commensurate with financial compensation in other recent MTurk studies (Buhrmester et al., 2011; Crump et al., 2013). Participants completed all study measures again during the second phase of data collection and were provided with increased compensation ($1.75).

The second period of data collection occurred approximately three months after the first, following precedent for assessing sensitivity to change as well as the emerging idea that psychological (in)flexibility may be more dynamic and state- or context-dependent than previously thought (Doorley et al., 2020; Stratford et al., 1996; Rolffs et al., 2018). Two-to-three-month time-lags are also common in studies evaluating the effects of ACT on psychological (in)flexibility, which provides further assurance of consistency and applicability of a three-month lag (Levin et al., 2020; Peltz et al., 2020). In order to assess sensitivity to change, four additional items were included in the assessment battery to identify a subset of participants who did not perceive a change in psychological (in)flexibility from Time 1 to Time 2 (Davidson & Keating, 2002; Shaw & Rogge, 2016). Specifically, participants responded to a general query (i.e., "How has your ability to be aware of the present moment, be open to a variety of experiences, and engage in behaviors that are self-directed and consistent with your values changed, if at all, in the months since you last completed this survey?") at the beginning of the survey. Additionally, three separate items, assessing three proposed dyads of the Hexaflex (i.e., openness to experience ["How has your ability to be open to a variety of experiences changed, if at all, in the months since you last completed this survey?"], present moment awareness ["How has your ability to be aware of the present moment changed, if at all, in the months since you last completed this survey?"], and values-directed action ["How has your ability to engage in behaviors that are self-directed and consistent with your values changed, if at all, in the months

since you last completed this survey?"]), were completed by participants. These three items were randomized among study measures across all participants. A single overall item was used to identify the subset of participants reporting that they did not experience a change in psychological (in)flexibility between Time 1 and Time 2 because all of the measures of interest are supposed to assess the same construct. Moreover, what is emphasized in the technical definitions that have been supplied for the construct that is assessed by each of the psychological (in)flexibility measures appears to be consistent across measures. The three additional questions were included in order to allow for additional investigation in differentiating perceived change across the components of the Hexaflex over time. Participants were instructed to rate their degree of change from Time 1 to Time 2 on a 7-point scale (-3 *got a lot worse* to 3 *got a lot better*) following the precedent set in previous studies examining sensitivity to change (Shaw & Rogge, 2016; Rolffs et al., 2018). The minimum detectable change was calculated for each measure of psychological (in)flexibility based on the responses of the subset of participants who report that they did not experience a change in psychological (in)flexibility between Time 1 and Time 2.

Measures

Psychological (In)Flexibility

Acceptance and Action Questionnaire-II. The Acceptance and Action Questionnaire-II (AAQ-II) is brief, commonly used measure of psychological (in)flexibility that provides a global score (Bond et al., 2011). Participants indicate how "true" each statement is for them by (e.g., "I'm afraid of my feelings") using a 7-point scale (1 = *never true* to 7 = *always true*). As was reviewed previously, evaluations of the psychometric properties of the AAQ-II have varied substantially. In general, researchers have demonstrated that the measure is reliable with a stable factor structure (Fledderus et al., 2012; Pennato et al., 2013; Ruiz et al., 2016). However, many

questions have been raised about the validity of the measure (Rochefort et al., 2018; Vaughan-Johnston et al., 2017; Wolgast, 2014). Internal consistency for the AAQ-II has ranged from acceptable to excellent in previous research (McAndrews et al., 2019). Internal consistency in the present sample was excellent (See Tables 1 and 2).

Acceptance and Action Questionnaire-3. The Acceptance and Action Questionnaire-3 (AAQ-3; Ong et al., 2019) is a 7-item modified version of the AAQ-II. The wording of most items was modified to reflect changes that would improve clarity and item functioning (Ong et al., 2020). Participants will indicate how "true" each statement is for them (e.g., "I'm so afraid of my feelings that I don't do things that I care about") by using a 7-point scale (1 = *never true* to 7 = *always true*). Given the novelty of the AAQ-3, its psychometric properties have not been extensively examined. It has demonstrated excellent internal consistency and good discriminant validity in student and community samples (Ong et al., 2020). Internal consistency in the present sample was excellent (See Tables 1 and 2).

Multidimensional Psychological Flexibility Inventory. The Multidimensional Psychological Flexibility Inventory (MPFI; Rolffs et al., 2018) is a 60-item measure of psychological flexibility and inflexibility. The scale consists of 12 subscales with six representing flexibility and six representing inflexibility. Participants were be asked rate each item (e.g., "When I had an upsetting thought or emotion, I tried to give it space rather than ignoring it" and "Negative feelings often trapped me in inaction") based on how true the item was for them in the past two weeks on a six-point scale (0 = *Never True* to 5 = *Always True*). Higher scores represent higher levels of the dimension being assessed. The scales and subscales of the MPFI have exhibited adequate internal consistency and evidence of convergent (e.g., AAQ-II, Avoidance and Fusion Questionnaire) and discriminant (e.g., emotional intelligence,

curiosity) validity (Rolffs et al., 2018). Further examination of the MPFI has provided additional evidence for the purposed factor structure as well as evidence that the MPFI has improved predictive validity relative to the AAQ-II (Rogge et al., 2019; Seidler et al., 2020). Internal consistency in the present sample was excellent for both the flexibility and inflexibility scale (See Tables 1 and 2).

Comprehensive Assessment of Acceptance and Commitment Therapy Processes. The Comprehensive Assessment of Acceptance and Commitment Therapy Processes (CompACT) is a 23-item measure that aligns with a conceptualization of the ACT Hexaflex in which psychological (in)flexibility is represented by three dyadic pairs (Francis et al., 2016; Hayes et al., 2011). Participants will be asked to rate each item (e.g., One of my big goals is to be free from painful emotions" and "Even when something is important to me, I'll rarely do it if there is a chance it will upset me") using a seven-point scale (0 = *Strongly Disagree* to 6 = *Strongly Agree*). Approximately half of the items are reverse scored, and higher scores on the CompACT reflect greater psychological flexibility. In the initial development and validation of the measure, the CompACT exhibited good convergent validity when compared to the AAQ-II and adequate discriminant validity when compared to a measure of social desirability (Francis et al. 2016). Further, the CompACT was correlated with clinical outcomes in a theory-consistent manner and items did not cross-load with items measuring general distress, suggesting that the CompACT successfully mitigated concerns related to discriminant validity. At this time, few additional psychometric evaluations of the CompACT have been conducted. One recent study did find that the CompACT accounted for greater variance in relevant clinical outcomes when compared to the AAQ-II (Rogge et al., 2019). Internal consistency in the present sample ranged from questionable to good (See Tables 1 and 2).

Personalized Psychological Flexibility Inventory. The Personalized Psychological Flexibility Inventory (PPFI) is a 19-item measure intended to capture aspects of psychological (in)flexibility within the context of valued goal pursuit in an idiographic manner (Kashdan et al., 2020). The measure consists of three subscales which are intended to capture avoidance-based goal strategies, acceptance-based strategies, and harnessing negative emotion as a goal-pursuit strategy. The PPFI also offers a total score representing overall psychological (in)flexibility, which is calculated by combining the three subscales while reverse scoring the avoidance subscale. Respondents were asked to identify an important goal and "select the rating that best describes [their] thoughts and feelings about this goal" for each item (e.g., "I avoid the most difficult goal-related tasks," "I accept the setbacks while pursuing this goal," "I find worrying helpful to solve goal-related problems") on a seven-point scale (1 = *Strongly Disagree* to 7 = *Strongly Agree*) while keeping their important goal in mind. The PPFI also includes four items designed to capture potential covariates related to the selected goal such as level of stress and goal centrality. During initial development, the PPFI demonstrated strong discriminant validity when compared to measures of negative emotions as well as when compared to existing measures of psychological (in)flexibility such as the AAQ-II (Kashdan et al., 2020). Further, the measure seems to converge in a theory-consistent manner with measures of well-being and adaptive functioning. At this time, there does not appear to be any independent evaluations of the psychometric properties of the PPFI. Internal consistency ranged from acceptable to good across the samples utilized during initial development of the measure (Kashdan et al., 2020). Internal consistency in the present sample ranged from poor to good (See Tables 1 and 2).

Criterion Measures

Depression Anxiety Stress Scales-21. The Depression, Anxiety, and Stress Scales-21 (DASS-21; Lovibond & Lovibond, 1995) were used as a measure of negative emotionality for the purposes of this study. Fourteen of the twenty-one items on the measure represent anxiety (e.g., "I was worried about situations in which I might panic and make a fool of myself") or depression (e.g., "I couldn't seem to experience any positive feelings at all"). Respondents will be asked to select the response that "indicates how much each statement applied to [them] over the past week" using a four-point scale (0 = *Did not apply to me at all* to 3 = *Applied to me very much, or most of the time*). The psychometric properties of the DASS-21 are well-established in the literature (Antony et al., 1998; Brown et al., 1997). Further, measurement invariance across clinical and nonclinical samples has been demonstrated, suggesting that the DASS-21 scores can be meaningfully used and compared across differing samples (Antony et al., 1998; Crawford & Henry, 2003). The DASS has been used as a measure of general distress in several recent studies evaluating the discriminant validity of measures of psychological (in)flexibility (Ong et al., 2020; Tyndall et al., 2019). Internal consistency in the present sample was excellent (See Tables 1 and 2).

Big Five Inventory Neuroticism Subscale. The Big Five Inventory (BFI) is a brief measure of personality that includes an eight-item subscale measuring trait neuroticism in adults (John et al., 1991; John et al., 2008). Participants rate the extent to which they agree or disagree with each item (e.g., "is depressed, blue" and "gets nervous easily") on a five-point scale (1 = *disagree strongly* to 5 = *agree strongly*). Higher scores on the BFI neuroticism subscale are indicative of more frequent negative emotional experiences (e.g., experiencing stress and nervousness). The BFI Neuroticism subscale is a well-validated and widely used measure of neuroticism (John & Srivastava, 1999). The BFI Neuroticism scale has previously been used a

criterion measure for establishing discriminant validity for the AAQ-II and MPFI and has demonstrated medium to large correlations with these measures of psychological (in)flexibility across several studies (Rochefort et al., 2018; Rolffs et al., 2018; Vaughan-Johnston et al., 2017). Internal consistency in the present sample ranged from questionable to good (see Tables 1 and 2).

Positive and Negative Affect Schedule. The Positive and Negative Affect Schedule (PANAS) is a 20-item self-report measure of state negative and positive affect, though negative affect will be the primary focus for this study (Watson et al., 1988). Participants will be asked to "indicate the extent [they] feel" each item (e.g., "excited" and "scared") in the past week on a five-point scale (1 = *very slightly or not at all* to 5 = *extremely*). The PANAS has consistently demonstrated strong psychometric properties including reliability, theoretically consistent convergent and discriminant validity, and invariance across demographic groups (Crawford & Henry, 2004; Watson et al., 1988). The PANAS has exhibited large correlations with the AAQ-II in previous evaluations of validity (Rochefort et al., 2018; Vaughan-Johnston et al., 2017; Wolgast, 2014). Internal consistency in the present sample was excellent (See Tables 1 and 2).

The Social Desirability Scale-17. The Social Desirability Scale (SDS-17; Stöber, 2001) is a measure of the degree to which individuals present themselves in a culturally appropriate manner. The SDS-17 is based on the Marlowe-Crowne Social Desirability Scale (MC-SDS; Crowne & Marlowe, 1960) and was intended to provide items with more current wording in order to mitigate concerns about participants potentially understanding item content differently depending on their age. The SDS-17 consists of 17 self-report items (e.g., "I occasionally speak badly of others behind their back") which are rated true or false by participants. The SDS-17 has demonstrated adequate psychometric properties including convergent validity with the original

MC-SDS as well as discriminant validity for extraversion, neuroticism, and psychoticism (Stöber, 2001). Measures of social desirability have been previously utilized in several studies evaluating the validity of measures of psychological (in)flexibility and have typically exhibited small magnitude correlations with such measures (Bond et al., 2011; Francis et al., 2016). Internal consistency in the present sample ranged from poor to acceptable (see Tables 1 and 2).

Toronto Empathy Questionnaire. The Toronto Empathy Questionnaire (TEQ) is a 16-item unidimensional measure of the ability to understand and respond to the emotions of others (Spreng et al., 2009). Participants will be asked to "rate how frequently you feel or act in the manner described" by selecting a response to each item (e.g., "I can tell when others are sad even when they do not say anything") from a five-point scale (0 = *Never* to 4 = *Always*). The TEQ has been validated in several diverse samples. Results suggest that the unidimensional structure of the measure is stable and that the TEQ relates to other measures of empathy as expected (Kourmousi et al., 2017; Spreng et al., 2009). Previous research has also demonstrated high internal consistency and reliability over time (Totan et al., 2012). The TEQ recently exhibited small, non-significant relationships with several measures of psychological (in)flexibility (Thompson et al., 2019). Internal consistency in the present sample ranged from acceptable to good (see Tables 1 and 2).

Data Analytic Plan

All analyses were conducted using IBM SPSS version 26 and MPlus Version 8.4 (Muthén & Muthén, 1998-2019). Internal consistency for each measure was evaluated based on McDonald's omega (ω), an alternative to the traditional Cronbach's alpha that is robust to violations of unidimensionality and tau-equivalence (Hayes & Coutts, 2020). McDonald's ω was calculated based on the factor loadings of a single-factor maximum likelihood confirmatory

factor analysis. Three commonly recommended fit statistics were used to evaluate the fit of each model: the comparative fit index (CFI), the Tucker-Lewis index (TLI), and the standardized root mean square residual (SRMR). The following criteria for determining adequate fit were utilized: CFI and TLI should be near .95 and SRMR should be less than .08 (Hu & Bentler, 1999). Additionally, internal consistency of the measures of psychological (in)flexibility (i.e., AAQ-II, AAQ-3, MPFI, CompACT, and PPFI) were more thoroughly assessed by an examination of inter-item and item-total correlations. Test-retest reliability for each psychological (in)flexibility scale was estimated by calculating intraclass correlation coefficients (ICC) and their 95% confidence intervals based on a single rating, absolute agreement, two-way mixed effects model (Koo & Li, 2017).

The primary goal of evaluating construct validity was assessed by examining Pearson correlations between total scores for the psychological (in)flexibility measures and each outcome. Construct validity was also assessed via contrast analysis, a procedure which allows for comparisons between theoretically predicted correlations and obtained correlations between the measures of interest (Westen & Rosenthal, 2003). This "quantifying construct validity" procedure is intended to reduce imprecise or subjective interpretations of correlations in examinations of construct validity (Furr, 2018). This procedure involves generating predicted correlations between measures of psychological (in)flexibility and the criterion measures outlined above based on the findings of existing research. Following data collection, the actual correlations between measures were calculated and the predicted and actual correlations were compared with two effect sizes ($r_{alerting\text{-}CV}$ and $r_{contrast\text{-}CV}$). These effect sizes can range from +1 to -1, with larger, more positive values representing convergence between predicted and actual correlations, or stronger construct validity (Westen & Rosenthal, 2003). More specifically, the

$r_{alerting\text{-}CV}$ effect size represents the degree to which correlations that are predicted to be high are high, while the $r_{contrast\text{-}CV}$ effect size represents the same information after correcting for intercorrelations among criterion variables and for the absolute level of correlations between the target measure and criterion measures (Furr, 2018).

The final aim of this study was to assess the sensitivity of each measure of psychological (in)flexibility to change. Specifically, the sensitivity of the measures of psychological (in)flexibility to changes over time after accounting for measurement error was assessed. With regard to responsiveness, minimally detectable differences were compared across all measures of psychological (in)flexibility. Differences between time one and time two scores on each measure were compared using paired t-tests. Additionally, the Minimal Detectible Change index (MDC_{95}) was calculated following the procedures outlined by Stratford et al. (1996) as cited by Rolffs et al. (2018). In order to replicate their procedures, the subset of participants reporting no perceived change in psychological (in)flexibility was analyzed with repeated measures ANOVAs in order to estimate the standard error of repeated measurement (SE_{RM}). The SE_{RM} for each scale was calculated using the equation SE_{RM} = SQRT(2*MSE), where MSE is the mean squared error from the repeated measures ANOVAs (Guyatt et al., 1987). Minimum detectable change was then calculated by multiplying the SE_{RM} by 1.96.

Results

Time One

Internal Consistency. Internal consistency was estimated using Cronbach's α and McDonald's ω. More specifically, McDonald's ω was estimated based on the factor loadings of a single-factor maximum likelihood confirmatory factor analysis. Estimates of internal consistency were generally high across measures of psychological (in)flexibility, with the exception of the

CompACT and the PPFI (see Table 1 for a summary of internal consistency statistics). Both the CompACT and PPFI had questionable internal consistency in the current sample. Further, fit for the unidimensional model used to estimate ω was poor for both the CompACT and PPFI, suggesting that the items did not fit well together in terms of representing a single underlying construct.

Item total correlations and interitem correlations were also examined for each measure of psychological (in)flexibility. Item total correlations provide information about how well individual items perform in terms of contributing information to the total score for a measure and interitem correlations capture how consistent the items are with one another. Item total correlations for all measures of psychological (in)flexibility were above .4, indicating very good item discrimination. Both versions of the AAQ exhibited notably high interitem correlations, with the interitem correlation for all items exceeding the .15 - .50 range recommended by Clark and Watson (1995). The magnitude of these correlations is indicative of significant redundancy in the items. The MPFI Flexibility and MPFI Inflexibility scales similarly had several items with particularly high interitem correlations; however, the average interitem correlation for both MPFI scales aligned more closely with recommendations, suggesting relatively less redundancy. Finally, the interitem correlations for the CompACT and the PPFI varied to a greater degree than the MPFI and both versions of the AAQ, thus suggesting that while some items were repetitive, others were largely unrelated.

Contrast Analysis (Convergent and Discriminant Validity). Descriptive statistics and bivariate correlations for time one ($N = 474$) are displayed in Tables 1 and 3. Correlations with criterion measures were examined and effect sizes were calculated to quantify the fit of each measure within the prespecified nomological network. Both the AAQ-II and AAQ-3 generally

correlated as predicted with other measures of psychological (in)flexibility, though the relationship between both versions of the AAQ and the PPFI was unexpectedly small and non-significant. Correlations with other criterion measures aligned with predictions more variably, and, while the magnitude of observed relationships varied slightly, the AAQ-II and AAQ-3 exhibited similar relationships with criterion measures. Specifically, the magnitude of the correlations between the AAQ measures and measures of anxiety, depression, negative affect, neuroticism, and empathy were large in size, whereas the magnitude of the correlations with social desirability were small. Overall, construct validity effect size correlations between the AAQ-II and criterion measures were high ($r_{\text{contrast-CV}} = .96$; $r_{\text{alerting-CV}} = .72$), with similar associations observed between the AAQ-3 and criterion measures ($r_{\text{contrast-CV}} = .97$; $r_{\text{alerting-CV}} = .72$; see Tables 4 and 5).

The MPFI Inflexibility scale generally correlated as predicted with measures of psychological (in)flexibility and outcome measures, though correlations with outcome measures were higher than was predicted (see Table 6). The relationships between the MPFI Inflexibility scale and the AAQ-II and AAQ-3 were large, aligning well with predictions. The relationship between the MPFI Inflexibility and CompACT was also large, while the correlation with the PPFI was small. With regard to criterion measures, the MPFI Inflexibility scale exhibited large correlations with measures of depression, anxiety, negative affective, and empathy, with all of these relationships being larger than predicted. The correlation between neuroticism and the MPFI Inflexibility scale was also large, but the magnitude of the correlation aligned more closely with the predicted relationship. As expected, the correlation with social desirability was small. Effect size correlations suggested good fit within the predicted nomological network ($r_{\text{contrast-CV}} = .94$; $r_{\text{alerting-CV}} = .68$). Observed correlations between the MPFI Flexibility scale and criterion

measures generally aligned with predicted correlations with regard to the magnitude of the effects, but not the direction of observed relationships (see Table 7). The MPFI Flexibility scale demonstrated a large, positive relationship with the MPFI Inflexibility scale, which again aligned with the predicted relationship in magnitude of the effect, but not the predicted direction. Observed correlations with the AAQ-II, AAQ-3, and PPFI were all positive, and large in magnitude, while the correlation between the MPFI Flexibility and CompACT was small. With regard to criterion measures, the MPFI Flexibility scale exhibited medium, positive correlations with measures of depression, anxiety, and negative affect and small, positive correlations with measures of neuroticism, social desirability, and empathy. It is important to highlight that the MPFI Flexibility scale unexpectedly exhibited positive relationships with most other measures of psychological (in)flexibility and with criterion measures. Additionally, the magnitude of the correlations between the MPFI Flexibility scale and other measures were similar, suggesting that the measure does not discriminate well between related but distinct constructs. Effect size correlations were indicative of poor fit within the hypothesized nomological network ($r_{\text{contrast-CV}}$ = .39; $r_{\text{alerting-CV}}$ = -.55).

Both the CompACT and PPFI varied significantly from predicted correlations (see Tables 8 and 9). With regard to the CompACT, medium to large correlations were observed with other measures of psychological (in)flexibility, which suggests differences in the underlying construct being measured. Additionally, the CompACT demonstrated only a small magnitude correlation with the MPFI Flexibility scale. Effects size correlations ($r_{\text{contrast-CV}}$ = .72; $r_{\text{alerting-CV}}$ = .62) indicated relatively poorer fit within the theoretically derived nomological network. Further, medium to large correlations with nearly all criterion measures suggest that the CompACT does not adequately differentiate between related, but distinct constructs. The PPFI did not align with

predictions and tended to exhibit small correlations with other measures of psychological (in)flexibility, with the exception of the CompACT and MPFI Flexibility scales, which exhibited medium to large correlations with the PPFI. Additionally, the PPFI demonstrated small correlations with all criterion measures, indicating that it was not related to outcomes as expected. Effect size correlations ($r_{\text{contrast-CV}} = .45$; $r_{\text{alerting-CV}} = -.63$) indicated that the PPFI did not fit well within the specified nomological network, suggesting that the measure may not be adequately capturing its proposed underlying construct.

Exploratory Analysis

Exploratory analyses were conducted in an attempt to make sense of the unexpected direction of the correlations between the MPFI Flexibility scale and criterion measures. Per the recommendations in the literature addressed above (e.g., Chmielewski and Kucker, 2020), the results of quality control questions were not initially utilized as a strategy for participant selection, nor were they used to ensure adequate attention to the survey. However, these types of questions were embedded in the study to allow for enhanced data cleaning, if needed. These questions included items such as "I have experienced a fatal heart attack while watching television" and "Please select 'Somewhat agree' if you are paying attention right now" (Oppenheimer et al., 2009; Paas et al., 2018). To ensure that data quality was not causing the unexpected results, the sample was restricted to include only participants who answered three of three quality control questions correctly. This reduction allowed for the examination of the subset of participants who were the most attentive to the survey. The data ($N = 101$) were then analyzed following the same procedure utilized with the larger Time One sample.

Internal Consistency. Estimates of internal consistency were generally high across measures of psychological (in)flexibility (alphas from .87 to .96; see Table 2). While the MPFI

scales, CompACT, and PPFI demonstrated good internal consistency based on both Cronbach's α and McDonald's ω, the fit for the unidimensional model used to estimate ω was poor for these scales, suggesting that the items did not fit well together in terms of representing a single underlying construct. Item total correlations and interitem correlations were also examined for each measure of psychological (in)flexibility. Item total correlations for the AAQ-II, AAQ-3, and MPFI-Flexibility scale were above .4, which indicated that scale items are good at discriminating between those with high and low levels of the construct of interest (i.e., psychological [in]flexibility). In contrast, the item total correlations for the MPFI Inflexibility scale, CompACT, and PPFI were more variable, suggesting that while some items discriminated well other items did not. Both versions of the AAQ exhibited notably high interitem correlations, with the interitem correlation for all items exceeding the .15 - .50 range recommended by Clark and Watson (1995). The magnitude of these correlations is indicative significant redundancy in the items. The MPFI Flexibility and MPFI Inflexibility scales similarly had several items with particularly high interitem correlations; however, the average interitem correlation for both MPFI scales aligned more closely with recommendations, suggesting relatively less redundancy. Finally, the interitem correlations for the CompACT and the PPFI varied to a greater degree than the MPFI and both versions of the AAQ, suggesting that while some items were repetitive, others were largely unrelated.

Contrast Analysis (Convergent and Discriminant Validity). Descriptive statistics and bivariate correlations for the Time 1 exploratory sample ($N = 101$) are displayed in Tables 2 and 10. Both the AAQ-II and AAQ-3 correlated as predicted with each other and the MPFI Inflexibility scale. Importantly, both versions of the AAQ exhibited correlations in the proposed direction with the MPFI Flexibility scale and the CompACT, though the magnitude of these

relationships was smaller than hypothesized. The observed correlation between both versions of the AAQ and the PPFI was negative and smaller than was predicted as well. Correlations with other criterion measures tended to align with predictions, though larger than predicted correlations with measures of depression, anxiety, and negative affect were observed. Further, while the magnitude of observed relationships varied slightly, the AAQ-II and AAQ-3 exhibited similar relationships with criterion measures, with the exception of social desirability. The SDS-17 exhibited a small positive correlation with the AAQ-II and a small negative correlation with the AAQ-3. The magnitude of the correlations between the AAQ measures and measures of depression, anxiety, neuroticism, and negative affect were large in size. Overall, construct validity effect size correlations between the AAQ-II and criterion measures were high ($r_{\text{contrast-CV}}$ = .96; $r_{\text{alerting-CV}}$ = .75), with similar associations observed between the AAQ-3 and criterion measures ($r_{\text{contrast-CV}}$ = .98; $r_{\text{alerting-CV}}$ = .75; see Tables 11 and 12).

The MPFI Inflexibility scale generally correlated as predicted with several measures of psychological (in)flexibility and outcome measures, though correlations with outcome measures were larger than predicted (see Table 13). The correlations between the MPFI Inflexibility scale and the AAQ-II and AAQ-3 were large, aligning well with predictions. The correlation between the MPFI Inflexibility and CompACT was also large and positive in direction, while the correlation with the PPFI was small in size and negative in direction. Notably, the relationship between the MPFI Inflexibility and Flexibility scales was smaller than expected. With regard to criterion measures, the MPFI Inflexibility scale exhibited large correlations with measures of depression, anxiety, neuroticism, and negative affective, with all of these relationships being larger than predicted. The correlation with social desirability was small, as expected. In contrast, the correlation with the measure of empathy was medium, which was larger than expected.

Effect size correlations suggested good fit within the predicted nomological network ($r_{\text{contrast-CV}}$ = .94; $r_{\text{alerting-CV}}$ = .74).

The MPFI Flexibility scale generally correlated as predicted with the CompACT and PPFI but observed correlations between the MPFI Flexibility scale and the other measures of psychological (in)flexibility were smaller than expected (see Table 14). With regard to criterion measures, the MPFI Flexibility scale exhibited small, negative correlations with measures of depression and negative affect as well as small, positive correlations with measures of anxiety, and social desirability. It exhibited a medium, negative correlation with a measure of neuroticism and a medium, positive relationship with a measure of empathy. Effect size correlations were indicative of relatively poorer fit within the hypothesized nomological network ($r_{\text{contrast-CV}}$ = .73; $r_{\text{alerting-CV}}$ = .92).

The CompACT generally correlated as predicted with the other measures of psychological (in)flexibility and criterion measures with regard to directionality, though the magnitude of the observed correlations varied from predictions (see Table 15). Specifically, the CompACT exhibited medium to large correlations with other measures of psychological (in)flexibility, but these relationships were somewhat smaller in magnitude than was predicted. Additionally, the correlation between the CompACT and the PPFI was positive, which does not align with predictions. With regard to criterion measures, the CompACT demonstrated medium to large correlations with nearly all measures, suggesting that the CompACT does not adequately differentiate between related but distinct constructs. Effect size correlations ($r_{\text{contrast-CV}}$ = .35; $r_{\text{alerting-CV}}$ = .18) indicated poor fit within the theoretically derived nomological network.

The observed correlations between the PPFI and the AAQ-II, AAQ-3, MPFI Inflexibility scale, and CompACT varied quite substantially from predictions with regard to both magnitude

and direction (see Table 16). The large, positive correlation between the PPFI and MPFI Flexibility scale was as predicted. The PPFI exhibited medium to large correlations with measures of depression, anxiety, neuroticism, negative affect, and empathy as well as a small magnitude correlation with the measure of social desirability though these correlations were not in the predicted direction. Effect size correlations ($r_{\text{contrast-CV}} = .82$; $r_{\text{alerting-CV}} = -.73$) indicated that the PPFI exhibited relatively poorer fit within the specified nomological network compared to the AAQ-II, AAQ-3, and MPFI Inflexibility scale, suggesting that the measure may not be adequately capturing its proposed underlying construct.

Time Two

Test-Retest Reliability. Descriptive statistics and bivariate correlations for Time 2 (N = 127) are displayed in Table 17. Test-retest reliability was estimated by calculating ICC estimates and their 95% confidence intervals based on an absolute agreement, two-way mixed effects model. ICC estimates for measures of psychological (in)flexibility varied across the two administrations (see Table 18). The AAQ-II (ICC = .81, 95% CI .74, .87), the AAQ-3 (ICC = .82, 95% CI .76, .87), and the MPFI Inflexibility scale (ICC = .79, 95% CI .71, .84) exhibited good reliability. Test-retest reliability for the MPFI Flexibility scale (ICC = .64, 95% CI .52, .73), the CompACT (ICC = .51, 95% CI .37, .63), and the PPFI (ICC = .50, 95% CI .36, .62) was moderate across the two administrations.

Sensitivity to Change. Differences between the first and second administration of each measure of psychological (in)flexibility were first assessed with a series of paired samples t-tests. As expected, there were no significant differences in scores over time for any of the measures (see Table 19). Next, a subset of participants (n = 28) who denied any change in psychological (in)flexibility were identified to isolate changes in scores that would represent measurement

error. Within this subset of participants, the amount of change required to establish a significant difference in scores across time was quantified by calculating the minimum detectible change index (MDC_{95}) based on the mean squared error over time from a repeated measures ANOVA. Practically, the MDC_{95} provides the amount of change necessary to observe statistically significant differences for a given scale. While these values cannot be compared across scales, considering the amount of change needed in terms of standard deviation aids in interpretation (see Table 20). The AAQ-3 ($MDC_{95} = 7.45$) exhibited greater sensitivity to change compared to the AAQ-II ($MDC_{95} = 9.28$). The amount of change needed on both versions of the AAQ was equivalent to less than one standard deviation of change. The MPFI Flexibility ($MDC_{95} = 48.11$) and Inflexibility ($MDC_{95} = 43.55$) scale each required approximately 1.5 standard deviations to detect significant change. In contrast to the other measures, the CompACT ($MDC_{95} = 51.80$) required change consistent with more than three standard deviations to detect significant differences. Similarly, the PPFI ($MDC_{95} = 20.69$) required approximately two standard deviations of change across time to detect statistically significant differences. Thus, these results provide evidence that the CompACT and PPFI both demonstrated relatively less responsiveness to change than the other measures of psychological (in)flexibility included in this study.

Subsets of participants reporting no change were also identified based on each of the items assessing three proposed dyads of the Hexaflex (present moment awareness [$n = 30$], openness to experience [$n = 28$], and values-directed action [$n = 32$]). In general, the pattern of required change across the measures of psychological (in)flexibility was similar, with the exception of the AAQ-II and AAQ-3 (see Table 21). With regard to these two measures, the AAQ-3 required less change overall but required more change for each of the three dyads, which may suggest less specificity with regard to the Hexaflex components. Although the differences

were minor, the MPFI Flexibility scale demonstrated slightly greater sensitivity to changes in openness to experience relative to the MPFI Inflexibility scale, while the latter demonstrated greater sensitivity to changes in present moment awareness and values-directed action. The CompACT demonstrated similar sensitivity across the three domains. The PPFI exhibited a similar pattern, though it was slightly less sensitive to changes in openness to experience.

Discussion

The overarching aim of this study was to complete a thorough evaluation of the psychometric properties of several measures of psychological (in)flexibility via the examination of a number of precise indicators of reliability and validity. Evaluation of these measures in the same sample allowed for comparisons to be made about the respective utility of each measure for use in future clinical and research endeavors. Some study hypotheses were supported, and results consistently supported the reliability and validity of multiple measures. Specifically, the AAQ-II, AAQ-3, and MPFI Inflexibility scales exhibited consistent evidence of strong reliability and validity. It is important to note that several of the measures of psychological (in)flexibility exhibited mixed evidence of adequate reliability across the various indicators used in this study, though this finding was tempered after additional data cleaning and exploratory analyses were conducted.

Although study hypotheses favored the measures with greater content coverage, both the AAQ-II and AAQ-3 consistently demonstrated excellent internal consistency, good reliability over time, and the best fit within the specified nomological network. However, both of these measures exhibited some redundancy in item content, which was particularly notable given the relatively short length of these measures. Moreover, it is important to note that although the AAQ-II and AAQ-3 exhibited the strongest fit within the prespecified nomological network,

both measures demonstrated correlations with several criterion measures that were nearly as large as their correlation with one another. These correlations suggest a lack of adequate discrimination between measures of similar constructs and criterion-related measures, a finding that aligns with patterns that have been consistently identified in the existing literature (Rochefort et al., 2018; Wolgast, 2014). Despite this lack of discrimination, the AAQ-II and AAQ-3 demonstrated the best fit in the network based on the results of the contrast analyses. It should be noted, however, that the basis for selecting criterion measures was largely based on existing studies in which the psychometric properties of the AAQ-II were evaluated, which may have allowed the AAQ-II and AAQ-3 to align with the nomological network to a greater degree in this study than the other measures of psychological (in)flexibility, which were developed within slightly different theoretical frameworks. With specific regard to the AAQ-3, results of this study suggest that the revised wording translated to equivalent or slightly improved measure performance. Given the practical length and widespread use of the AAQ-II, it could be beneficial to consider further, more substantial revisions to this measure in future studies. More specifically, both versions of the AAQ demonstrated notably high item redundancy, which could be one focus of potential revisions.

One aspect of test construction that was not fully addressed by this study was the idea of construct clarity. As was previously discussed, the AAQ-II has been referred to as both a measure of psychological (in)flexibility and experiential avoidance, a potential flaw that several of the newer measures of psychological (in)flexibility intended to address (Kashdan et al., 2020; Rolffs et al., 2018). The results of this study suggest that both versions of the AAQ align well with other measures of psychological (in)flexibility. However, the specific components of psychological (in)flexibility, including experiential avoidance, were not considered in this study

due to power constraints and availability of measures. It would be beneficial in future studies to utilize a different network of criterion measures to assess how well the AAQ and other measures of psychological (in)flexibility align with the broader theoretical model of the construct. Further, a preliminary investigation of the factor structure of the MPFI suggests that the domain-specific factors may be redundant with the overarching general factors (i.e., flexibility and inflexibility) and do not provide incremental value beyond the general factors in predicting criterion-related variables (Thomas et al., 2021). Thus, further investigation of the domain-specific factors of the psychological (in)flexibility constructs seems pertinent.

The MPFI Flexibility and Inflexibility scales diverged somewhat in terms of their reliability and validity, especially with regard to relationships with criterion measures. The MPFI Inflexibility scale demonstrated fit within the nomological network, though like the AAQ, tended to correlate highly with criterion measures of anxiety, depression, and negative affect. In contrast, the MPFI Flexibility scale exhibited unexpected positive correlations with all criterion measures, the MPFI Inflexibility scale, and other (in)flexibility measures. This is particularly problematic because, in theory, the MPFI Flexibility scale should exhibit negative correlations with all of these measures. Results of a follow-up exploratory analysis provided more theoretically consistent results in that the MPFI Flexibility scale demonstrated negative correlations with the majority of the criterion measures. However, the magnitude of the correlation between the MPFI Flexibility and Inflexibility scales in the exploratory analysis was smaller than expected based on previous research (Rolffs et al., 2018). This finding is consistent with previous literature suggesting that psychological flexibility and inflexibility may be distinct constructs that differently relate to outcomes (Kashdan & Rottenberg, 2010; Rolffs, et al., 2018;

Stabbe et al., 2019). However, this discrepancy may also be a function, to some degree, of data that is not of the highest quality (see limitations and future directions).

In contrast with hypotheses, both the CompACT and PPFI demonstrated mixed evidence of reliability and validity across the indicators used in this study. These two measures demonstrated greater variability across the initial and exploratory analyses. Additionally, with specific regard to the CompACT, while the authors of the measure classified initial reliability as good, they relied on only inter-item correlations to demonstrate reliability and did not provide any other indicators of internal consistency (e.g., Cronbach's alpha) for the measure, making comparisons from the existing literature more difficult (Francis et al., 2016). The CompACT demonstrated moderate to good internal consistency based on the indicators of reliability utilized in this study. With regard to validity, the CompACT demonstrated relatively poorer alignment with predictions compared to the AAQ measures and the MPFI Inflexibility scale. Notably, the CompACT did not appear to have good discrimination between measures as it exhibited medium to large correlations with nearly all other measures. The PPFI demonstrated poor to good reliability across indicators and analyses. Regarding validity, the PPFI did not fit well within the prespecified nomological network, though fit improved in the exploratory analysis. The PPFI exhibited unexpected negative correlations with criterion measures across both the original and exploratory analyses.

An additional aim of this study was to evaluate the reliability of each measure of psychological (in)flexibility across time. This was an important aim as few studies to date have evaluated this aspect of measurement with regard to this construct, despite the fact that several of these measures are being used to evaluate outcomes across time in treatment studies (e.g., Levin et al., 2020; Meyer et al., 2018). Following the guidelines offered by Koo and Li (2017), the test-

retest reliability of the measures of psychological (in)flexibility ranged from moderate (PPFI, CompACT, MPFI Flexibility) to good (MPFI Inflexibility, AAQ-II, AAQ-3). However, it is important to consider the context of the construct when interpreting these statistics. Psychological (in)flexibility is thought to be relatively stable over time, though some previous research favors the idea that the construct may be more dynamic and state- or context-dependent than previously thought (Cherry et al., 2021; Doorley et al., 2020; Rolffs et al., 2018). Based on this idea, it is possible that the ICC estimates observed in this study may underestimate the reliability of the measures, particularly those that ask participants to focus on shorter time periods or specific contexts such as the MPFI and PPFI. Additionally, the AAQ measures, which conceptualize psychological (in)flexibility as a more stable construct exhibited greater stability over time.

The final aim of the study was to evaluate the sensitivity to change of each measure of psychological (in)flexibility. Sensitivity to change has been evaluated less often in this literature; however, this analysis is important because it offers insight into how much change would be required to detect significant differences across time. Given that changes in psychological (in)flexibility are considered the main theoretical driver of symptom changes in the ACT model, it is vital to be able to confidently identify such changes when evaluating the treatment model and its efficacy. Sensitivity to change in this study was quantified with the MDC_{95}. Unlike many of the indices utilized in this project, these estimates are not directly comparable across measures; therefore, results were interpreted in terms of standard deviations for each individual scale. Sensitivity to change varied substantially across the measures. Both versions of the AAQ exhibited greater sensitivity, requiring less than one standard deviation of change to detect significant differences. The CompACT and PPFI exhibited less sensitivity as both measures

required substantial change (i.e., more than two standard deviations) to detect differences across time.

This pattern of results was also generally observed across the three theoretical dyads: present moment awareness, openness to experience, and values-directed action. However, sensitivity of individual scales varied somewhat across the dyads. Specifically, the AAQ scales, MPFI scales, and PPFI were all less sensitive to openness to experience, while the CompACT demonstrated relatively less sensitivity to present moment awareness. These differences were relative, but, along with differing relationships to criterion measures, may speak to a larger pattern in which these different measures of psychological (in)flexibility capture the specific domains that make up the psychological (in)flexibility construct to different degrees. This analysis may serve as a helpful starting place for future evaluations of sensitivity to change.

Limitations and Future Directions

It is important to consider the results of this study within the context of potential limitations. First, the constructs selected for inclusion in the pre-specified nomological network may not have been optimal to provide a complete picture of validity. Specifically, the criterion measures selected for this study were primarily selected due to their relevance to criticisms levied against the AAQ-II (Rochefort et al., 2018; Wolgast, 2014). While constructs such as negative affect and neuroticism are relevant to establishing discriminant validity for measures of psychological (in)flexibility, they represent only a piece of the larger nomological network. For those measures that have subscales based on the ACT Hexaflex, an alternative approach to selecting criterion measures could be to choose measures that align with the constructs comprising psychological (in)flexibility such as acceptance, present moment awareness, and defusion. These comparisons could provide a better sense of how well these measures align with

the underlying theoretical constructs they propose to capture. Further, given that a particularly notable finding of this study was the large correlations between several of the measures of psychological (in)flexibility and measures of depression and anxiety, an additional consideration for the nomological network could be to include more detailed or nuanced measures of these outcomes including specific symptoms. A recent validation of the MPFI scales utilized outcome measures that provided information about well-being more generally such as fatigue and social engagement, which may also be an appropriate option for studies in the general population rather than a clinical sample (Grégoire et al., 2020). Overall, it seems important to diversify and expand the constructs within the nomological network in order to learn more about the validity of the target measures.

Another prominent limitation of this study was poor participant retention across time. After accounting for individuals excluded for poor effort or attention, approximately 27% of Time 1 participants were retained at Time 2. Several barriers may have impeded retention across time, but the clearest barrier was an inability to directly contact potential participants. A significant proportion (approximately 50%) of participants no longer had active Mturk accounts at Time 2. The method used to contact participants did not allow individuals without active accounts to be directly contacted in order to maintain anonymity. As such, they could see the study as available if they re-activated their account at a later time, but they could not receive email reminders. While Mturk is a reliable source of data that provides a reasonably accurate representation of the general population (Chandler & Shapiro, 2016), it may not be optimal for future longitudinal studies due to the high rate of participant turnover. Future studies examining the test-retest reliability or sensitivity to change over time of measures of psychological (in)flexibility may benefit from recruiting participants from a different source.

Participants also appeared to exhibit notable declines in attention across the Time 1 survey. This decline may have been related to the way measures were presented, with the MPFI, a 60-item measure, being presented on one page rather than being broken up across several pages. While this is only one hypothesis, the MPFI scales demonstrated particularly unusual patterns of results from the Time 1 data. Additional data cleaning and exploratory analyses were used to address this limitation. The results of the updated analyses did favor the idea that inattentive responding was a problem as the directionality of the observed relationships was largely as hypothesized following the additional data cleaning. Unfortunately, these analyses significantly reduced the size of the sample and several unexpected findings remained, the most prominent of which was a notably small correlation between the MPFI Flexibility and Inflexibility scales. Replicating this study with greater consideration of how the measures are presented would likely be important to improve confidence in the findings.

Finally, it is important to note that the majority of the sample across this study identified as white and college educated. Approximately 82% of the participants comprising the sample were white and college educated at Time 1 and approximately 79% at Time 2. As a result of the lack of population representation in this sample, the results of this study may not generalize to the population-at-large. This is an important consideration as previous literature has suggested significant bias toward highly educated, western samples in psychological research (Henrich et al., 2010; Thalmayer et al., 2021). Despite some research in supporting the ability to recruit diverse samples from Mturk (e.g., Behrend et al., 2011; Chandler & Shapiro, 2016; Hauser & Schwarz, 2016), it seems possible that the strategy for data collection, which relied on access to internet across time, may have systematically excluded participants from lower socioeconomic

backgrounds. It will be important for this set of research questions to be evaluated in more diverse, representative samples in the future.

Conclusion

Despite significant limitations, this study represents an important step in understanding the reliability and validity of a variety of measures of psychological (in)flexibility. Rather than relying on traditional methods of estimating the psychometric properties of the measures of interest, this study utilized a wider variety of indicators, based on advances in the psychometric literature, to increase precision in the evaluation of these vital test characteristics. Results from this study suggest that psychometric properties vary substantially across measures of psychological (in)flexibility, with the AAQ-II, AAQ-3, and MPFI Inflexibility scale demonstrating the clearest evidence of reliability and validity across the indicators examined and within the nomological network specified for this study. It is important to note that while several measures aligned well with the network, evidence suggested that all of the measures of psychological (in)flexibility exhibited problems with discriminant validity. Specifically, the correlations between psychological (in)flexibility and symptom measures tended to be either very large or very small rather than aligning with predictions. These findings offer numerous potential directions for future research including refinement of existing measures, further exploration of the reliability and construct validity of the measures, and tests of alternative versions of the theoretical network in which psychological (in)flexibility is embedded. While the results of this study must be considered within the context of study limitations, the methods of the study represent a step forward in refining the measurement of psychological (in)flexibility, and, in turn, improving understanding of the construct more generally.

References

Antony, M. M., Bieling, P. J., Cox, B. J., Enns, M. W., & Swinson, R. P. (1998). Psychometric properties of the 42-item and 21-item versions of the Depression Anxiety Stress Scales in clinical groups and a community sample. *Psychological Assessment, 10*(2), 176–181.

A-tjak, J. G., Davis, M. L., Morina, N., Powers, M. B., Smits, J. A., & Emmelkamp, P. M. (2015). A meta-analysis of the efficacy of acceptance and commitment therapy for clinically relevant mental and physical health problems. *Psychotherapy and Psychosomatics, 84(1),* 30-36.

Arch, J. J., Eifert, G. H., Davies, C., Vilardaga, J. C. P., Rose, R. D., & Craske, M. G. (2012). Randomized clinical trial of cognitive behavioral therapy (CBT) versus acceptance and commitment therapy (ACT) for mixed anxiety disorders. *Journal of Consulting and Clinical Psychology, 80(5),* 750-765.

Bai, Z., Luo, S., Zhang, L., Wu, S., & Chi, I. (2020). Acceptance and commitment therapy (ACT) to reduce depression: a systematic review and meta-analysis. *Journal of Affective Disorders, 260,* 728-737.

Barlow, D. H., Allen, L. B., & Choate, M. L. (2004). Toward a unified treatment for emotional disorders. *Behavior Therapy, 35*(2), 205-230.

Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior Research Methods, 43,* 800-813.

Benfer, N., Spitzer, E. G., & Bardeen, J. R. (2021). Efficacy of third wave cognitive behavioral therapies in the treatment of posttraumatic stress: A meta-analytic study. *Journal of Anxiety Disorders,* Article 102360.

Benoy, C., Knitter, B., Knellwolf, L., Doering, S., Klotsche, J., & Gloster, A. T. (2018). Assessing psychological flexibility: validation of the open and engaged state questionnaire. *Journal of Contextual Behavioral Science, 12,* 253-260.

Benoy, C., Knitter, B., Schumann, I., Bader, K., Walter, M., & Gloster, A. T. (2019). Treatment sensitivity: Its importance in the measurement of psychological flexibility. *Journal of Contextual Behavioral Science, 13,* 121-125.

Biglan, A., & Hayes, S. C. (1996). Should the behavioral sciences become more pragmatic? The case for functional contextualism in research on human behavior. *Applied and Preventive Psychology: Current Scientific Perspectives, 5,* 47–57.

Bond, F.W., & Bunce, D. (2003). The role of acceptance and job control in mental health, job satisfaction, and work performance. *Journal of Applied Psychology, 88,* 1057–1067.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy, 42,* 676-688.

Bramwell, K., & Richardson, T. (2018). Improvements in depression and mental health after acceptance and commitment therapy are related to changes in defusion and values-based action. *Journal of Contemporary Psychotherapy, 48*(1), 9-14.

Brown, T. A., Chorpita, B. F., Korotitsch, W., & Barlow, D. H. (1997). Psychometric properties of the Depression Anxiety Stress Scales (DASS) in clinical samples. *Behaviour Research and Therapy, 35*(1), 79-89.

Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality data? *Perspectives on Psychological Science, 6,* 3-5.

Buhrmester, M. D., Talaifar, S., & Gosling, S. D. (2018). An evaluation of Amazon's Mechanical Turk, its rapid rise, and its effective use. *Perspectives on Psychological Science, 13*(2), 149-154.

Carretero-Dios, H., Eid, M., & Ruch, W. (2011). Analyzing multitrait-mulitmethod data with multilevel confirmatory factor analysis: An application to the validation of the State-Trait Cheerfulness Inventory. *Journal of Research in Personality, 45*(2), 153-164.

Chandler, J., & Shapiro, D. (2016). Conducting clinical research using crowdsourced convenience samples. *Annual Review of Clinical Psychology, 12,* 53-81.

Chawla, N., & Ostafin, B. (2007). Experiential avoidance as a functional dimensional approach to psychopathology: An empirical review. *Journal of Clinical Psychology, 63*(9), 871-890.

Cherry, K. M., Vander Hoeven, E., Patterson, T. S., & Lumley, M. N. (2021). Defining and measuring "psychological flexibility": A narrative scoping review of diverse flexibility and rigidity constructs and perspectives. *Clinical Psychology Review,* Article 101973.

Chmielewski, M., & Kucker, S. C. (2020). An Mturk crisis? Shifts in data quality and the impact on study results. *Social Psychological and Personality Science, 11*(4), 464-473.

Clark, L. A., & Watson, D. (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment, 31,* 1412-1427.

Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine, 74,* 327–337.

Crawford, J. R., & Henry, J. D. (2003). The Depression Anxiety Stress Scales (DASS): Normative data and latent structure in a large non-clinical sample. *British Journal of Clinical Psychology, 42*(2), 111-131.

Crawford, J. R., & Henry, J. D. (2004). The Positive and Negative Affect Schedule (PANAS): Construct validity, measurement properties and normative data in a large non-clinical sample. *British Journal of Clinical Psychology, 43*(3), 245-265.

Crump, M. J., McDonnell, J. V., & Gureckis, T. M. (2013). Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research. *PLOS ONE, 8,* Article e57410.

Curran, P. G. (2016). Methods for the detection of carelessly invalid responses in survey data. *Journal of Experimental Social Psychology, 66,* 4–19.

Daly, T. M., & Nataraajan, R. (2015). Swapping bricks for clicks: Crowdsourcing longitudinal data on Amazon Turk. *Journal of Business Research, 68,* 2603-2609.

Davis, R. N., & Nolen-Hoeksema, S. (2000). Cognitive inflexibility among ruminators and nonruminators. *Cognitive Therapy and Research, 24,* 699-711.

Davidson, M., & Keating, J. L. (2002). A comparison of five low back disability questionnaires: reliability and responsiveness. *Physical Therapy, 82*(1), 8-24.

Deacon, B. J., Fawzy, T. I., Lickel, J. J., & Wolitzky-Taylor, K. B. (2011). Cognitive defusion versus cognitive restructuring in the treatment of negative self-referential thoughts: An investigation of process and outcome. *Journal of Cognitive Psychotherapy, 25,* 218-232.

Dimidjian, S., Arch, J. J., Schneider, R., Desormeau, P., Felder, J. N., & Segal, Z. V. (2016). Considering meta-analysis, meaning, and metaphor: A systematic review and critical examination of "third wave" cognitive and behavioral therapies. *Behavior Therapy, 47*(6), 886–905.

Doi, S., Ito, M., Takebayashi, Y., Muramatsu, K., & Horikoshi, M. (2018). Factorial validity and invariance of the 7-item generalized anxiety disorder scale (GAD-7) among populations

with and without self-reported psychiatric diagnostic status. *Frontiers in Psychology, 9,* Article 1741.

Doorley, J. D., Goodman, F. R., Kelso, K. C., & Kashdan, T. B. (2020). Psychological flexibility: What we know, what we do not know, and what we think we know. *Social and Personality Psychology Compass, 14*(12), 1-11.

Fergus, T. A. (2015). I really believe I suffer from a health problem: examining an association between cognitive fusion and healthy anxiety. *Journal of Clinical Psychology, 71,* 920-934.

Firsching, V. J., Villanueva, J., Rinner, T. B., Benoy, C., Kuhweide, V., Brogli, S., & Gloster, A. T. (2018). Measuring psychological flexibility in a context sensitive manner – development and preliminary psychometric properties of a short and accessible questionnaire. *Poster-Vorstellung anlässlich der ACBS World Conference 16, Montreal, Canada.*

Fledderus, M., Oude Voshaar, M. A., ten Klooster, P. M., & Bohlmeijer, E. T. (2012). Further evaluation of the psychometric properties of the Acceptance and Action Questionnaire–II. *Psychological Assessment, 24,* 925-936.

Fok, C. C. T., & Henry, D. (2015). Increasing the sensitivity of measures to change. *Prevention Science, 16*(7), 978-986.

Forman, E. M., Chapman, J. E., Herbert, J. D., Goetter, E. M., Yuen, E. K., & Moitra, E. (2012). Using session-by-session measurement to compare mechanisms of action for acceptance and commitment therapy and cognitive therapy. *Behavior Therapy, 43,* 341-354.

Forman, E. M., Herbert, J. D., Moitra, E., Yeomans, P. D., & Geller, P. A. (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification, 31,* 772-799.

Foster, S. L., & Cone, J. D. (1995). Validity issues in clinical assessment. *Psychological Assessment, 7*(3), 248–260.

Francis, A. W., Dawson, D. L., & Golijani-Moghaddam, N. (2016). The development and validation of the Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT). *Journal of Contextual Behavioral Science, 5,* 134-145.

Furr, R. M. (2018). *Psychometrics: An introduction*. (3rd ed.). SAGE publications.

Furr, R. M., & Heuckeroth, S. (2019). The "Quantifying Construct Validity" procedure: Its role, value, interpretations, and computation. *Assessment, 26,* 555-566.

Gámez, W., Chmielewski, M., Kotov, R., Ruggero, C., & Watson, C. (2011). Development of a measure of experiential avoidance: The Multidimensional Experiential Avoidance Questionnaire. *Psychological Assessment, 23,* 692–713.

Gloster, A. T., Klotsche, J., Ciarrochi, J., Eifert, G., Sonntag, R., Wittchen, H. U., & Hoyer, J. (2017). Increasing valued behaviors precedes reduction in suffering: Findings from a randomized controlled trial using ACT. *Behaviour Research and Therapy, 91,* 64-71.

Gloster, A. T., Walder, N., Levin, M. E., Twohig, M. P., Karekla, M. (2020). The empirical status of acceptance and commitment therapy: A review of meta-analyses. *Journal of Contextual Behavioral Science, 18,* 181-192.

Godbee, M., & Kangas, M. (2020). The Relationship Between Flexible Perspective Taking and Emotional Well-Being: A Systematic Review of the "Self-As-Context" Component of Acceptance and Commitment Therapy. *Behavior Therapy, 51,* 917-932.

Grégoire, S., Gagnon, J., Lachance, L., Shankland, R., Dionne, F., Kotsou, I., … & Rogge, R. D. (2020). Validation of the English and French versions of the multidimensional psychological flexibility inventory short form (MPFI-24). *Journal of Contextual Behavioral Science, 18,* 99-110.

Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin, 103*(2), 265–275.

Guyatt, G. H., Walter, S., & Norman, G. (1987). Measuring change over time: Assessing the *usefulness of evaluative instruments. Journal of Chronic Disease, 40,* 171-178.

Hacker, T., Stone, P., & MacBeth, A. (2016). Acceptance and commitment therapy–do we know enough? Cumulative and sequential meta-analyses of randomized controlled trials. *Journal of Affective Disorders, 190,* 551-565.

Hauser, D. J. & Schwarz, N. (2016). Attentive turkers: Mturk participants perform better on online attention checks than do subject pool participants. *Behavior Research Methods, 48,* 400-407.

Hays, R. D., & Hadorn, D. (1992). Responsiveness to change: An aspect of validity, not a separate dimension. *Quality of Life Research, 1,* 73-75.

Hayes, A. F., & Coutts, J. J. (2020). Use Omega rather than Cronbach's Alpha for estimating reliability, but…. *Communication Methods and Measures, 14*(1), 1-24.

Hayes, S. C. (1989). *Rule-governed behavior: Cognition, contingencies, and instructional control.* New York, NY: Plenum Press.

Hayes, S. C. (1993). Analytic goals and the varieties of scientific contextualism. In S. C. Hayes, L. J. Hayes, H. W. Reese, & T. R. Sarbin (Eds.), *Varieties of scientific contextualism* (pp. 11–27). Reno, NV: Context Press.

Hayes, S. C. (2004). Acceptance and commitment therapy, relational frame theory, and the third wave of behavioral and cognitive therapies. *Behavior Therapy, 35,* 639-665.

Hayes, S., C, Barnes-Holmes, D, & Roche, B. (Eds.). (2001). *Relational frame theory: A post-Skinnerian account of human language and cognition.* New York: Kluwer Academic/Plenum.

Hayes, S. C., Barnes-Holmes, D., & Wilson, K. G. (2012). Contextual behavioral science: Creating a science more adequate to the challenge of the human condition. *Journal of Contextual Behavioral Science, 1,* 1-16.

Hayes, S. C., Hayes, L. J., Reese, H. W., & Sarbin, T. R. (1993). Analytic goals and the varieties of scientific contextualism. In S. C. Hayes, L. J. Reese, & T. R. Sarbin (Eds.), *Varieties of scientific contextualism* (pp. 11–27). Reno, NV: Context Press.

Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: Model, processes, and outcomes. *Behavior Research and Therapy, 44,* 1-25.

Hayes, S. C., Pistorello, J., & Levin, M. E. (2012). Acceptance and commitment therapy as a unified model of behavior change. *The Counseling Psychologist, 40,* 976-1002.

Hayes, S. C., Strosahl, K. D., & Wilson, K. G. (2012). *Acceptance and commitment therapy: The process and practice of mindful change.* Guilford Press.

Hayes, S. C., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., . . . McCurry, S. M. (2004). Measuring experiential avoidance: A preliminary test of a working model. *Psychological Record, 54,* 553-578.

Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, aware, and active: Contextual approaches as an emerging trend in the behavioral and cognitive therapies. *Annual Review of Clinical Psychology, 7*(4), 141–168.

Hayes, S. C., Wilson, K. G., Gifford, E. V., Follette, V. M., & Strosahl, K. (1996). Experiential avoidance and behavioral disorders: A functional dimensional approach to diagnosis and treatment. *Journal of Consulting and Clinical Psychology, 64,* 1152-1168.

Henrich, J., Heine, S. J., & Norenzayan, A. (2010). Most people are not WEIRD. *Nature, 466* (7302), 29.

Hofmann, S. G., Heering, S., Sawyer, A. T., & Asnaani, A. (2009). How to handle anxiety: The effects of reappraisal, acceptance, and suppression strategies on anxious arousal. *Behaviour Research and Therapy, 47,* 389-394.

John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.

John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (pp. 114-158). New York, NY: Guilford Press.

John, O. P., & Srivastava, S. (1999). *The Big-Five trait taxonomy: History, measurement, and theoretical perspectives* (Vol. 2, pp. 102-138). Berkeley: University of California.

Johnson, R. E., Rosen, C. C., Chang, C., Djurdjevic, E., & Taing, M. U. (2012). Recommendations for improving the construct clarity of higher-order multidimensional constructs. *Human Resource Management Review, 22,* 62-72.

Kashdan, T. B., Disabato, D. J., Goodman, F. R., Doorley, J. D., & McKnight, P. E. (2020). Understanding psychological flexibility: A multimethod exploration of pursuing valued goals despite the presence of distress. *Psychological Assessment, 32,* 829-850.

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review, 30,* 865-878.

Keith, M. G., Tay, L., & Harms, P. D. (2017). Systems perspective of Amazon Mechanical Turk for organizational research: Review and recommendations. *Frontiers in Psychology, 8,* Article 1359.

Koo, T. K., & Li, M. Y. (2017). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine, 15*(2), 155-163.

Kourmousi, N., Amanaki, E., Tzavara, C., Merakou, K., Barbouni, A., & Koutras, V. (2017). The Toronto Empathy Questionnaire: reliability and validity in a nationwide sample of Greek teachers. *Social Sciences, 6*(2), 1-14.

Krafft, J., Haeger, J. A., & Levin, M. E. (2019). Comparing cognitive fusion and cognitive reappraisal as predictors of college student mental health. *Cognitive Behaviour Therapy, 48*(3), 241-252.

Levin, M. E., Krafft, J., & Twohig, M. P. (2020). Examining processes of change in an online acceptance and commitment therapy dismantling trial with distressed college students. *Journal of Contextual Behavioral Science, 17,* 10-16.

Levin, M. E., Krafft, J., Hicks, E. T., Pierce, B., & Twohig, M. P. (2020). A randomized dismantling trial of the open and engaged components of acceptance and commitment therapy in an online intervention for distressed college students. *Behaviour Research and Therapy, 126,* Article 103557.

Levin, M. E., Krafft, J., Pierce, B., & Potts, S. (2018). When is experiential avoidance harmful in the moment? Examining global experiential avoidance as a moderator. *Journal of Behavior Therapy and Experimental Psychiatry, 61,* 158-163.

Levin, M. E., MacLane, C., Daflos, S., Seeley, J. R., Hayes, S. C., Biglan, A., & Pistorello, J. (2014). Examining psychological inflexibility as a transdiagnostic process across psychological disorders. *Journal of Contextual Behavioral Science, 3(3),* 155-163.

Lovibond, P. F., & Lovibond, S. H. (1995). The structure of negative emotional states: Comparison of the Depression Anxiety Stress Scales (DASS) with the Beck Depression and Anxiety Inventories. *Behaviour Research and Therapy, 33,* 335-343.

Luoma, J. B., Sabucedo, P., Eriksson, J., Gates, N., & Pilecki, B. C. (2019). Toward a contextual psychedelic-assisted therapy: Perspectives from Acceptance and Commitment Therapy and contextual behavioral science. *Journal of Contextual Behavioral Science, 14,* 136-145.

Masuda, A., Twohig, M. P., Stormo, A. R., Feinstein, A. B., Chou, Y. Y., & Wendell, J. W. (2010). The effects of cognitive defusion and thought distraction on emotional discomfort and believability of negative self-referential thoughts. *Journal of Behavior Therapy and Experimental Psychiatry, 41*(1), 11-17.

McAndrews, Z., Richardson, J., & Stopa, L. (2019). Psychometric properties of acceptance measures: A systematic review. *Journal of Contextual Behavioral Science, 12,* 261-277.

McCracken, L. M., & Gutiérrez-Martínez, O. (2011). Processes of change in psychological flexibility in an interdisciplinary group-based treatment for chronic pain based on Acceptance and Commitment Therapy. *Behaviour Research and Therapy, 49,* 267-274.

Meyer, E. C., Walser, R., Hermann, B., Bash, H., DeBeer, B. B., Morissette, S. B., Kimbrel, N. A., Kwok, O., Batten, S. V., & Schnurr, P. P. (2018). Acceptance and commitment therapy for co-occurring posttraumatic stress disorder and alcohol use disorders in veterans: Pilot treatment outcomes. Journal of Traumatic Stress, 31(5), 781–789.

Muthén, L. K., & Muthén, B. O. (2019). MPlus (Version 8.4) [Computer software]. Los Angeles, CA: Muthén & Muthén.Ong, C. W., Pierce, B. G., Woods, D. W., Twohig, M. P., & Levin, M. E. (2019). The Acceptance and Action Questionnaire–II: An item response theory analysis. *Journal of Psychopathology and Behavioral Assessment, 41*(1), 123-134.

Ong, C. W., Pierce, B. G., Petersen, J. M., Barney, J. L., Fruge, J. E., Levin, M. E., & Twohig, M. P. (2020). A psychometric comparison of psychological inflexibility measures: Discriminant validity and item performance. *Journal of Contextual Behavioral Science, 18,* 34-47.

Ong, C. W., Pierce, B. G., Woods, D. W., Twohig, M. P., & Levin, M. E. (2019). The Acceptance and Action Questionnaire–II: An item response theory analysis. *Journal of Psychopathology and Behavioral Assessment, 41*(1), 123-134.

Oppenheimer, D. M., Meyvis, T., & Davidenko, N. (2009). Instructional manipulation checks: Detecting satisficing to increase statistical power. *Journal of Experimental Social Psychology, 45*(4), 867-872.

Öst, L. G. (2014). The efficacy of acceptance and commitment therapy: an updated systematic review and meta-analysis. *Behaviour Research and Therapy, 61,* 105-121.

Paas, L. J., Dolnicar, S., & Karlsson, L. (2018). Instructional manipulation checks: A longitudinal analysis with implications for MTurk. *International Journal of Research in Marketing, 35*(2), 258-269.

Paolacci, G., Chandler, J., & Ipeirotis, P. G. (2010). Running experiments on amazon mechanical turk. *Judgment and Decision making, 5,* 411-419.

Peer, E., Vosgerau, J., & Acquisti, A. (2014). Reputation as a sufficient condition for data quality on Amazon Mechanical Turk. *Behavior Research Methods, 46,* 1023–1031.

Peltz, J. S., Rogge, R. D., Bodenlos, J. S., Kingery, J. N., & Pigeon, W. R. (2020). Changes in psychological inflexibility as a potential mediator of longitudinal links between college students' sleep problems and depressive symptoms. *Journal of Contextual Behavioral Science, 15,* 110-122.

Pennato, T., Berrocal, C., Bernini, O., & Rivas, T. (2013). Italian version of the Acceptance and Action Questionnaire-II (AAQ-II): Dimensionality, reliability, convergent and criterion validity. *Journal of Psychopathology and Behavioral Assessment, 35*(4), 552-563.

Petri, J. M., Weathers, F. W., Witte, T. K., & Silverstein, M. W. (2020). The Detailed Assessment of Posttraumatic Stress–Second Edition (DAPS-2): Initial Psychometric Evaluation in an MTurk-Recruited, Trauma-Exposed Community Sample. *Assessment, 27*(6), 1116-1127.

Pilecki, B. C., & McKay, D. (2012). An experimental investigation of cognitive defusion. *The Psychological Record, 62*(1), 19-40.

Rochefort, C., Baldwin, A. S., & Chmielewski, M. (2018). Experiential avoidance: An examination of the construct validity of the AAQ-II and the MEAQ. *Behavior Therapy, 49,* 435-449.

Rogge, R. D., Daks, J. S., Dubler, B. A., & Saint, K. J. (2019). It's all about the process: Examining the convergent validity, conceptual coverage, unique predictive validity, and

clinical utility of ACT process measures. *Journal of Contextual Behavioral Science, 14,* 90-102.

Rolffs, J. L., Rogge, R. D., & Wilson, K. G. (2018). Disentangling components of flexibility via the Hexaflex model: Development and validation of the Multidimensional Psychological Flexibility Inventory (MPFI). *Assessment, 25,* 458-482.

Ruiz, F. J., Suárez-Falcón, J. C., Cárdenas-Sierra, S., Durán, Y., Guerrero, K., & Riaño-Hernández, D. (2016). Psychometric properties of the Acceptance and Action Questionnaire–II in Colombia. *The Psychological Record, 66*(3), 429-437.

Seidler, D., Stone, B., Clark, B. E., Koran, J., & Drake, C. E. (2020). Evaluating the factor structure of the Multidimensional Psychological Flexibility Inventory: An independent replication and extension. *Journal of Contextual Behavioral Science, 17,* 23-31.

Serowik, K. L., Khan, A. J., LoCurto, J., & Orsillo, S. M. (2018). The conceptualization and measurement of values: A review of the psychometric properties of measures developed to inform values work with adults. *Journal of Psychopathology and Behavioral Assessment, 40*(4), 615-635.

Shapiro, D. N., Chandler, J., & Mueller, P. A. (2013). Using Mechanical Turk to study clinical populations. *Clinical Psychological Science, 1*(2), 213-220.

Shaw, A. M., & Rogge, R. D. (2016). Evaluating and refining the construct of sexual quality with item response theory: Development of the Quality of Sex Inventory. *Archives of Sexual Behavior, 45*(2), 249-270.

Spreng, R. N., McKinnon, M. C., Mar, R. A., & Levine, B. (2009). The Toronto Empathy Questionnaire: Scale development and initial validation of a factor-analytic solution to multiple empathy measures. *Journal of Personality Assessment, 91*(1), 62-71.

Stabbe, O. K., Rolffs, J. L., & Rogge, R. D. (2019). Flexibly and/or inflexibly embracing life: Identifying fundamental approaches to life with latent profile analyses on the dimensions of the Hexaflex model. *Journal of Contextual Behavioral Science, 12,* 106-118.

Stöber, J. (2001). The Social Desirability Scale-17 (SDS-17): Convergent validity, discriminant validity, and relationship with age. *European Journal of Psychological Assessment, 17*(3)*,* 222-232.

Stockton, D., Kellett, S., Berrios, R., Sirois, F., Wilkinson, N., & Miles, G. (2019). Identifying the underlying mechanisms of change during acceptance and commitment therapy (ACT): A systematic review of contemporary mediation studies. *Behavioural and Cognitive Psychotherapy, 47,* 332-362.

Stratford, P. W., Binkley, J. M., & Riddle, D. L. (1996). Health status measures: Strategies and analytic methods for assessing change scores. *Physical Therapy, 76,* 1109-1123.

Stratford, P. W., Binkley, J., Solomon, P., Finch, E., Gill, C., & Moreland, J. (1996). Defining the minimum level of detectable change for the Roland-Morris questionnaire. *Physical Therapy, 76,* 359-365.

Thalmayer, A. G., Toscanelli, C., & Arnett, J. J. (2021). The neglected 95% revisited: Is American psychology becoming less American? *American Psychologist, 76* (1), 116–129.

Thomas, K. N., Bardeen, J. R., Witte, T. K., Rogers, T. A., Benfer, N., & Clauss, K. (2021). An Examination of the Factor Structure of the Multidimensional Psychological Flexibility Inventory. *Assessment,* 1-16.

Thompson, M., Bond, F. W., & Lloyd, J. (2019). Preliminary psychometric properties of the Everyday Psychological Inflexibility Checklist. *Journal of Contextual Behavioral Science, 12,* 243-252.

Totan, T., Dogan, T., & Sapmaz, F. (2012). The Toronto Empathy Questionnaire: Evaluation of Psychometric Properties among Turkish University Students. *Eurasian Journal of Educational Research, 46,* 179-198.

Trindade, I. A., & Ferreira, C. (2014). The impact of body image-related cognitive fusion on eating psychopathology. *Eating Behaviors, 15*(1), 72-75.

Twohig, M. P. (2012). Introduction: The basics of acceptance and commitment therapy. *Cognitive and Behavioral Practice, 19,* 499-507.

Twohig, M. P., & Levin, M. E. (2017). Acceptance and commitment therapy as a treatment for anxiety and depression: A review. *Psychiatric Clinics, 40*(4), 751-770.

Tyndall, I., Waldeck, D., Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2019). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science, 12,* 278-284.

Vaughan-Johnston, T. I., Quickert, R. E., & MacDonald, T. K. (2017). Psychological flexibility under fire: Testing the incremental validity of experiential avoidance. *Personality and Individual Differences, 105,* 335-349.

Vermeersch, D. A., Lambert, M. J., & Burlingame, G. M. (2000). Outcome questionnaire: Item sensitivity to change. *Journal of Personality Assessment, 74,* 242-261.

Villatte, J. L., Vilardaga, R., Villatte, M., Vilardaga, J. C. P., Atkins, D. C., & Hayes, S. C. (2016). Acceptance and Commitment Therapy modules: Differential impact on treatment processes and outcomes. *Behaviour Research and Therapy, 77,* 52-61.

Villatte, M., Villatte, J. L., & Hayes, S. C. (2016). *Mastering the clinical conversation: Language as intervention.* Guilford Press.

Vowles, K. E., McCracken, L. M., & O'Brien, J. Z. (2011). Acceptance and values-based action in chronic pain: a three-year follow-up analysis of treatment effectiveness and process. *Behaviour Research and Therapy, 49,* 748-755.

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.

Wenze, S. J., Gaugler, T. L., Sheets, E. S., & DeCicco, J. M. (2018). Momentary experiential avoidance: Within-person correlates, antecedents, and consequences and between-person moderators. *Behaviour Research and Therapy, 107,* 42-52.

Westen, D., & Rosenthal, R. (2003). Quantifying construct validity: Two simple measures. *Journal of Personality and Social Psychology, 84,* 608-618.

Williams, M. T., Ching, T. H., Tellawi, G., Siev, J., Dowell, J., Schlaudt, V., ... & Wetterneck, C. T. (2018). Assessing sexual orientation symptoms in obsessive-compulsive disorder: Development and validation of the Sexual Orientation Obsessions and Reactions Test (SORT). *Behavior Therapy, 49*(5), 715-729.

Wilson, K. G., Sandoz, E. K., Kitchens, J., & Roberts, M. (2010). The Valued Living Questionnaire: Defining and measuring valued action within a behavioral framework. *The Psychological Record, 60,* 249-272.

Wolgast, M. (2014). What does the Acceptance and Action Questionnaire (AAQ-II) really measure? *Behavior Therapy, 45,* 831-839.

Wood, D., Harms, P. D., Lowman, G. H., & DeSimone, J. A. (2017). Response speed and response consistency as mutually validating indicators of data quality in online samples. *Social Psychological and Personality Science, 8,* 454–464.

Wright, A. G., & Woods, W. C. (2020). Personalized models of psychopathology. *Annual Review of Clinical Psychology, 16,* 49-74.

Wright, A. G., & Zimmermann, J. (2019). Applied ambulatory assessment: Integrating idiographic and nomothetic principles of measurement. *Psychological Assessment, 31,* 1467-1480.

Yarrish, C., Groshon, L., Mitchell, J. D., Appelbaum, A., Klock, S., Winternitz, T., & Friedman-Wheeler, D. G. (2019). Finding the signal in the noise: Minimizing responses from bots and inattentive humans in online research. *The Behavior Therapist, 42,* 235-242.

Zhou, H., & Fishbach, A. (2016). The pitfall of experimenting on the web: How unattended selective attrition leads to surprising (yet false) research conclusions. *Journal of Personality and Social Psychology, 111*(4), 1-12.

Zvolensky, M. J., Feldner, M. T., Leen-Feldner, E., & Yartz, A. R. (2005). Exploring basic processes underlying acceptance and mindfulness. In S. Orsillo & L. Roemer (Eds.), *Acceptance and mindfulness-based approaches to anxiety: Conceptualization and treatment* (pp. 325–359). New York: Springer.

Table 1

Descriptive Statistics and Internal Consistency at Time 1

	M (SD)	Cronbach's α	McDonald's ω	CFI	TLI	SRMR	Item Total Correlations	Interitem Correlations	Mean Interitem Correlation
AAQ-II	31.44 (9.70)	.93	.94	.994	.991	.016	.734 - .810	.620 - .750	.67
AAQ-3	30.63 (10.10)	.94	.94	.968	.952	.026	.709 - .835	.551 - .736	.68
MPFI Flexibility	122.92 (25.18)	.96	.96	.901	.894	.041	.488 - .721	.252 - .607	.44
MPFI Inflexibility	113.38 (32.13)	.97	.97	.907	.901	.047	.492 - .822	.296 - .730	.54
CompACT	71.32 (11.47)	.61	.89	.572	.530	.132	.541 - .731	.160 - .675	.45
PPFI	66.69 (8.58)	.58	.57	.711	.663	.105	.518 - .658	.191 - .726	.41
DASS Depression	10.67 (5.60)	.91	.90	.981	.971	.027	-	-	-
DASS Anxiety	10.53 (5.46)	.91	.90	.968	.953	.032	-	-	-
BFI-N	25.53 (5.21)	.68	.77	.924	.894	.065	-	-	-
PANAS Negative Affect	29.25 (10.44)	.94	.94	.968	.959	.030	-	-	-
SDS-17	9.28 (2.37)	.41	-	-	-	-	-	-	-
TEQ	35.40 (7.67)	.71	.72	.548	.479	.14	-	-	-

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 2

Descriptive Statistics and Internal Consistency at Time 1 (N = 101)

	M (SD)	Cronbach's α	McDonald's ω	CFI	TLI	SRMR	Item Total Correlations	Interitem Correlations	Mean Interitem Correlation
AAQ-II	21.89 (12.08)	.96	.96	.93	.90	.03	.798 - .880	.662 - .887	.77
AAQ-3	19.85 (11.66)	.95	.95	.96	.94	.03	.724 - .892	.567 - .854	.74
MPFI Flexibility	114.36 (30.82)	.97	.97	.74	.72	.08	.409 - .813	.078 - .691	.52
MPFI Inflexibility	80.37 (32.24)	.97	.97	.66	.63	.11	.234 - .844	-.109 - .852	.51
CompACT	80.08 (19.60)	.88	.88	.51	.46	.18	.095 - .744	-.324 - .760	.24
PPFI	67.03 (14.61)	.84	.81	.44	.34	.21	.294 - .630	-.177 - .830	.26
DASS Depression	5.18 (5.92)	.94	.94	.97	.96	.03	-	-	-
DASS Anxiety	4.61 (5.11)	.92	.92	.96	.94	.03	-	-	-
BFI-N	19.58 (7.45)	.86	.86	.85	.78	.09	-	-	-
PANAS Negative Affect	18.82 (8.48)	.92	.92	.85	.81	.06	-	-	-
SDS-17	9.84 (3.37)	.73	-	-	-	-	-	-	-
TEQ	44.40 (9.95)	.86	.82	.47	.39	.21	-	-	-

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 3

Bivariate Correlations for Time 1

	1	2	3	4	5	6	7	8	9	10	11	12
1. AAQ-II	-											
2. AAQ-3	.89***	-										
3. MPFI Flexibility	.44***	.42***	-									
4. MPFI Inflexibility	.86***	.84***	.57***	-								
5. CompACT	-.50***	-.52***	.12**	-.48***	-							
6. PPFI	-.05	-.07	.48***	-.01	.40***	-						
7. DASS Depression	.78***	.82***	.41***	.79***	-.50***	-.06	-					
8. DASS Anxiety	.77***	.81***	.47***	.77***	-.43***	.02	.87***	-				
9. BFI Neuroticism	.65***	.67***	.12**	.62***	-.51***	-.22***	.69***	.61***	-			
10. PANAS Negative Affect	.72***	.78***	.39***	.74***	-.43***	-.03	.77***	.76***	.66***	-		
11. SDS-17	-.12**	-.16**	.04	-.15**	.12**	.15**	-.20***	-.14**	-.28***	-.19***	-	
12. TEQ	-.50***	-.56***	.01	-.50**	.48***	.17***	-.56***	-.57***	-.42***	-.53***	.23***	-

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 4

Predicted and Observed Correlations Between AAQ-II and Criterion Measures, Raw λs, and Integer Values of λ

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	AAQ-II	Predicted *r*	Raw λs	Raw λs as integers
AAQ-3	.89***	.80	0.50	5
MPFI Flexibility	.44***	-.50	-0.80	-8
MPFI Inflexibility	.86***	.80	0.5	5
CompACT	-.50***	-.80	-1.10	-11
PPFI	-.05	.80	0.50	5
DASS Depression	.78***	.60	0.30	3
DASS Anxiety	.77***	.60	0.30	3
BFI Neuroticism	.65***	.60	0.30	3
PANAS Negative Affect	.72***	.60	0.30	3
SDS-17	-.12**	-.10	-0.4	-4
TEQ	-.50***	-.10	-0.4	-4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 5

Predicted and Observed Correlations Between AAQ-3 and Criterion Measures, Raw λs, and Integer Values of λ

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	AAQ-3	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	.89***	0.80	0.50	5
MPFI Flexibility	.42***	-0.50	-0.80	-8
MPFI Inflexibility	.84***	0.80	0.5	5
CompACT	-.52***	-0.80	-1.10	-11
PPFI	-.07	0.80	0.50	5
DASS Depression	.82***	0.60	0.30	3
DASS Anxiety	.81***	0.60	0.30	3
BFI Neuroticism	.67***	0.60	0.30	3
PANAS Negative Affect	.78***	0.60	0.30	3
SDS-17	-.16**	-0.10	-0.4	-4
TEQ	-.56***	-0.10	-0.4	-4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 6

Predicted and Observed Correlations Between MPFI Inflexibility and Criterion Measures, Raw λs, and Integer Values of λ

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	MPFI Inflexibility	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	.86***	0.80	0.50	5
AAQ-3	.84***	0.80	0.50	5
MPFI Flexibility	.57***	-0.50	-0.8	-8
CompACT	-.48***	-0.80	-1.10	-11
PPFI	-.01	0.80	0.50	5
DASS Depression	.79***	0.60	0.30	3
DASS Anxiety	.77***	0.60	0.30	3
BFI Neuroticism	.62***	0.60	0.30	3
PANAS Negative Affect	.74***	0.60	0.30	3
SDS-17	-.15**	-0.10	-0.4	-4
TEQ	-.50**	-0.10	-0.4	-4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 7

Predicted and Observed Correlations Between MPFI Flexibility and Criterion Measures, Raw λs, and Integer Values of λ

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	MPFI Flexibility	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	.44***	-0.50	-0.36	-4
AAQ-3	.42***	-0.50	-0.36	-4
MPFI Inflexibility	.57***	-0.50	-0.36	-4
CompACT	.12**	0.50	0.64	6
PPFI	.48***	0.50	0.64	6
DASS Depression	.41***	-0.40	-0.26	-3
DASS Anxiety	.47***	-0.40	-0.26	-3
BFI Neuroticism	.12**	-0.40	-0.26	-3
PANAS Negative Affect	.39***	-0.40	-0.26	-3
SDS-17	.04	0.30	0.44	4
TEQ	.01	0.30	0.44	4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 8

Predicted and Observed Correlations Between CompACT and Criterion Measures, Raw λs, and Integer Values of λ

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	CompACT	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	-.50***	-0.80	-0.35	-4
AAQ-3	-.52***	-0.80	-0.35	-4
MPFI Flexibility	.12**	0.50	0.95	9
MPFI Inflexibility	-.48***	-0.80	-0.35	-4
PPFI	.40***	-0.80	-0.35	-4
DASS Depression	-.50***	-0.60	-0.15	-2
DASS Anxiety	-.43***	-0.60	-0.15	-2
BFI Neuroticism	-.51***	-0.60	-0.15	-2
PANAS Negative Affect	-.43***	-0.60	-0.15	-2
SDS-17	.12**	0.10	0.55	5
TEQ	.48***	0.10	0.55	5

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 9

Predicted and Observed Correlations Between PPFI and Criterion Measures, Raw λs, and Integer Values of λ

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	PPFI	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	-.05	0.80	0.41	4
AAQ-3	-.07	0.80	0.41	4
MPFI Flexibility	.48***	0.50	0.11	1
MPFI Inflexibility	-.01	0.80	0.41	4
CompACT	.40***	-0.80	-1.19	-12
DASS Depression	-.06	0.60	0.21	2
DASS Anxiety	.02	0.60	0.21	2
BFI Neuroticism	-.22***	0.60	0.21	2
PANAS Negative Affect	-.03	0.60	0.21	2
SDS-17	.15**	-0.10	-0.49	-5
TEQ	.17***	-0.10	-0.49	-5

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 10

Bivariate Correlations for Time 1 (N = 101)

	1	2	3	4	5	6	7	8	9	10	11	12
1. AAQ-II	-											
2. AAQ-3	.88***	-										
3. MPFI Flexibility	-.12	-.20*	-									
4. MPFI Inflexibility	.79***	.76***	-.02	-								
5. CompACT	-.54***	-.57***	.44***	-.60***	-							
6. PPFI	-.32**	-.35***	.56***	-.35***	.57***	-						
7. DASS Depression	.80***	.86***	-.14	.75***	-.54***	-.31**	-					
8. DASS Anxiety	.73***	.77***	.03	.66***	-.40***	-.20*	.85***	-				
9. BFI Neuroticism	.64***	.67***	-.28**	.64***	-.52***	-.39***	.70***	.50***	-			
10. PANAS Negative Affect	.76***	.80***	-.12	.72***	-.53***	-.28**	.80***	.77***	.61***	-		
11. SDS-17	.01	-.09	.17	-.05	.04	.19	-.21*	-.06	-.25*	-.11	-	
12. TEQ	-.34**	-.47***	.36***	-.29**	.32**	.25*	-.42***	-.50***	-.13	-.41***	.16	-

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 11

Predicted and Observed Correlations Between AAQ-II and Criterion Measures, Raw λs, and Integer Values of λ (N = 101)

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	AAQ-II	Predicted *r*	Raw λs	Raw λs as integers
AAQ-3	.88***	.80	0.50	5
MPFI Flexibility	-.12	-.50	-0.80	-8
MPFI Inflexibility	.79***	.80	0.5	5
CompACT	-.54***	-.80	-1.10	-11
PPFI	-.32**	.80	0.50	5
DASS Depression	.80***	.60	0.30	3
DASS Anxiety	.73***	.60	0.30	3
BFI Neuroticism	.64***	.60	0.30	3
PANAS Negative Affect	.76***	.60	0.30	3
SDS-17	.01	-.10	-0.4	-4
TEQ	-.34**	-.10	-0.4	-4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 12

Predicted and Observed Correlations Between AAQ-3 and Criterion Measures, Raw λs, and Integer Values of λ (N = 101)

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	AAQ-3	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	.88***	0.80	0.50	5
MPFI Flexibility	-.20*	-0.50	-0.80	-8
MPFI Inflexibility	.76***	0.80	0.5	5
CompACT	-.57***	-0.80	-1.10	-11
PPFI	-.35***	0.80	0.50	5
DASS Depression	.86***	0.60	0.30	3
DASS Anxiety	.77***	0.60	0.30	3
BFI Neuroticism	.67***	0.60	0.30	3
PANAS Negative Affect	.80***	0.60	0.30	3
SDS-17	-.09	-0.10	-0.4	-4
TEQ	-.47***	-0.10	-0.4	-4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 13

Predicted and Observed Correlations Between MPFI Inflexibility and Criterion Measures, Raw λs, and Integer Values of λ (N = 101)

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	MPFI Inflexibility	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	.79***	0.80	0.50	5
AAQ-3	.76***	0.80	0.50	5
MPFI Flexibility	-.02	-0.50	-0.8	-8
CompACT	-.60***	-0.80	-1.10	-11
PPFI	-.35***	0.80	0.50	5
DASS Depression	.75***	0.60	0.30	3
DASS Anxiety	.66***	0.60	0.30	3
BFI Neuroticism	.64***	0.60	0.30	3
PANAS Negative Affect	.72***	0.60	0.30	3
SDS-17	-.05	-0.10	-0.4	-4
TEQ	-.29**	-0.10	-0.4	-4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 14

Predicted and Observed Correlations Between MPFI Flexibility and Criterion Measures, Raw λs, and Integer Values of λ (N = 101)

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	MPFI Flexibility	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	-12	-0.50	-0.36	-4
AAQ-3	-.20*	-0.50	-0.36	-4
MPFI Inflexibility	-.02	-0.50	-0.36	-4
CompACT	.44***	0.50	0.64	6
PPFI	.56***	0.50	0.64	6
DASS Depression	-.14	-0.40	-0.26	-3
DASS Anxiety	.03	-0.40	-0.26	-3
BFI Neuroticism	-.28**	-0.40	-0.26	-3
PANAS Negative Affect	-.12	-0.40	-0.26	-3
SDS-17	.17	0.30	0.44	4
TEQ	.36***	0.30	0.44	4

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 15

Predicted and Observed Correlations Between CompACT and Criterion Measures, Raw λs, and Integer Values of λ (N = 101)

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	CompACT	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	-.54***	-0.80	-0.35	-4
AAQ-3	-.57***	-0.80	-0.35	-4
MPFI Flexibility	.44***	0.50	0.95	9
MPFI Inflexibility	-.60***	-0.80	-0.35	-4
PPFI	.57***	-0.80	-0.35	-4
DASS Depression	-.54***	-0.60	-0.15	-2
DASS Anxiety	-.40***	-0.60	-0.15	-2
BFI Neuroticism	-.52***	-0.60	-0.15	-2
PANAS Negative Affect	-.53***	-0.60	-0.15	-2
SDS-17	.04	0.10	0.55	5
TEQ	.32**	0.10	0.55	5

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 16

Predicted and Observed Correlations Between PPFI and Criterion Measures, Raw λs, and Integer Values of λ (N = 101)

	Observed Correlations	Predicted Correlations and λs		
Criterion Variable	PPFI	Predicted *r*	Raw λs	Raw λs as integers
AAQ-II	-.32***	0.80	0.41	4
AAQ-3	-.35***	0.80	0.41	4
MPFI Flexibility	.56***	0.50	0.11	1
MPFI Inflexibility	-.35***	0.80	0.41	4
CompACT	.57***	-0.80	-1.19	-12
DASS Depression	-.54***	0.60	0.21	2
DASS Anxiety	-.40***	0.60	0.21	2
BFI Neuroticism	-.52***	0.60	0.21	2
PANAS Negative Affect	-.53***	0.60	0.21	2
SDS-17	.04	-0.10	-0.49	-5
TEQ	.32**	-0.10	-0.49	-5

*$p < .05$, **$p < .01$, ***$p < .001$

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 17

Descriptive Statistics for Time 2

	M (SD)	1	2	3	4	5	6	7	8	9	10	11	12
1. AAQ-II	28.46 (12.03)	-											
2. AAQ-3	27.32 (12.24)	.94***	-										
3. MPFI Flexibility	124.102 (25.95)	.13	.12	-									
4. MPFI Inflexibility	104.57 (36.57)	.92***	.91***	-.24**	-								
5. CompACT	75.72 (16.92)	-.74***	-.74***	.24**	-.73***	-							
6. PPFI	67.13 (9.36)	-.28**	-.29**	.54***	-.21**	.43***	-						
7. DASS Depression	8.87 (6.42)	.85***	.85***	.14	.85***	-.68***	-.33***	-					
8. DASS Anxiety	8.32 (6.17)	.81***	.82***	.24**	.84***	-.63***	-.14	.88***	-				
9. BFI Neuroticism	22.52 (6.31)	.70***	.71***	-.27**	.65***	-.65***	-.44***	.67***	.56***	-			
10. PANAS Negative Affect	25.70 (11.04)	.80***	.83***	.18*	.82***	-.60***	-.23**	.82***	.82***	.65***	-		
11. SDS-17	9.40 (2.96)	-.22**	-.24**	.26**	-.20**	.24**	.23**	-.26**	-.24**	-.33***	-.24**	-	
12. TEQ	36.94 (10.01)	-.54***	-.54***	.14	-.57***	.55***	.17*	-.52***	-.56***	-.44***	-.50***	.34***	-

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory; DASS = Depression Anxiety and Stress Scales; BFI = Big Five Inventory; PANAS = Positive and Negative Affect Schedule; SDS-17 = Social Desirability Scale-17; TEQ = Toronto Empathy Questionnaire.

Table 18

Intraclass Correlation Coefficients

	Intraclass Correlation Coefficient	95% Confidence Interval	
		Lower Bound	Upper Bound
AAQ-II	.81	.74	.87
AAQ-3	.82	.76	.87
MPFI Flexibility	.64	.52	.73
MPFI Inflexibility	.79	.71	.84
CompACT	.51	.37	.63
PPFI	.50	.36	.62

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory.

Table 19

Paired Samples T-Tests

	Mean	Standard Deviation	*t* value	*df*	Sig (two-tailed)
Time 1 AAQ-II	28.51	11.72	.085	126	.93
Time 2 AAQ-II	28.46	12.03			
Time 1 AAQ-3	27.95	11.97	.999	126	.32
Time 2 AAQ-3	27.31	12.24			
Time 1 MPFI Flexibility	122.72	26.69	-.696	126	.49
Time 2 MPFI Flexibility	124.10	25.95			
Time 1 MPFI Inflexibility	105.56	34.92	.477	126	.64
Time 2 MPFI Inflexibility	104.57	36.57			
Time 1 CompACT	73.13	15.78	-1.811	126	.07
Time 2 CompACT	75.72	16.92			
Time 1 PPFI	68.42	10.25	1.491	126	.14
Time 2 PPFI	67.13	9.36			

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory

Table 20

Minimum Detectible Change Index for Measures of Psychological (In)flexibility

	MDC_{95}	SD Units	Possible Range of Scores
AAQ-II	9.28	.77	7-49
AAQ-3	7.45	.61	7-49
MPFI Flexibility	48.11	1.62	30-180
MPFI Inflexibility	43.55	1.37	30-180
CompACT	51.80	3.06	0-138
PPFI	20.69	2.21	15-105

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory

Table 21

Minimum Detectible Change Index for Measures of Psychological (In)flexibility Based on Hexaflex Dyads

	MDC_{95}	SD Units
Present Moment Awareness (*n* = 30)		
AAQ-II	8.32	.68
AAQ-3	9.50	.76
MPFI Flexibility	44.54	1.33
MPFI Inflexibility	42.17	1.27
CompACT	52.97	2.17
PPFI	20.68	1.46
Openness to Experience (*n* = 28)		
AAQ-II	8.77	.80
AAQ-3	11.98	1.12
MPFI Flexibility	46.74	1.53
MPFI Inflexibility	52.94	1.70
CompACT	50.79	2.08
PPFI	19.93	1.81
Values-Directed Action (*n* = 32)		
AAQ-II	8.56	.68
AAQ-3	10.34	.81
MPFI Flexibility	45.08	1.41
MPFI Inflexibility	44.39	1.25
CompACT	50.34	2.07
PPFI	21.32	1.54

Note. AAQ = Acceptance and Action Questionnaire; MPFI = Multidimensional Psychological Flexibility Inventory; CompACT = Comprehensive Assessment of Acceptance and Commitment Therapy Processes; PPFI = Personalized Psychological Flexibility Inventory