



HHS Public Access

Author manuscript

Behav Ther. Author manuscript; available in PMC 2024 November 01.

Published in final edited form as:

Behav Ther. 2023 November ; 54(6): 971–988. doi:10.1016/j.beth.2022.07.005.

Acceptance and Commitment Therapy Processes and Mediation: Challenges and How to Address Them

Joanna J. Arch,

University of Colorado Boulder and University of Colorado Cancer Center

Joel N. Fishbein,

University of Colorado Boulder

Lauren B. Finkelstein,

University of Colorado Boulder

Jason B. Luoma

Portland Psychotherapy Clinic, Research and Training Center

Abstract

Acceptance and commitment therapy (ACT) emphasizes a focus on theory-driven processes and mediating variables, a laudable approach. The implementation of this approach would be advanced by addressing five challenges, including (a) distinguishing ACT processes in measurement contexts, (b) developing and rigorously validating measures of ACT processes, (c) the wide use of psychometrically weaker ACT process measures and the more limited use of stronger measures in earlier work, (d) the inconsistency of past evidence that ACT processes are sensitive or specific to ACT or mediate ACT outcomes specifically, and (e) improving statistical power and transparency. Drawing on the existing literature, we characterize and provide evidence for each of these challenges. We then offer detailed recommendations for how to address each challenge in ongoing and future work. Given ACT's core focus on theorized processes, improving the measurement and evaluation of these processes would significantly advance the field's understanding of ACT.

Keywords

acceptance and commitment therapy; mediation; measurement; assessment; replication

Acceptance and commitment therapy (ACT; Hayes et al., 1999, 2012) is a highly creative, flexible, and generative intervention model that has been evaluated in over 900 randomized trials to date (Hayes, 2022). These trials address an extremely broad range of psychological and behavioral problems and disorders, reflecting ACT's ability to leverage its core intervention strategies, perspectives, and skills to address diverse forms of human struggle and suffering. The ACT model, also known as the psychological flexibility model, posits six overlapping therapeutic processes: acceptance, cognitive defusion, contact with the present

Address correspondence to Joanna Arch, Ph.D., University of Colorado Boulder, Department of Psychology and Neuroscience, 345 UCB Muenzinger, Boulder, CO 80309-0345. Joanna.Arch@Colorado.edu.

The authors declare no other conflicts of interest.

moment, self-as-context, values, and committed action (Hayes et al., 2006), which together are known as the Hexaflex model. A widely cited claim is that the targeted processes¹ and mechanisms of the ACT Hexaflex model are theory driven, clearly defined, and account for ACT's benefits (Hayes et al., 2013). Strong measurement of ACT process variables remains core to ACT and to contextual behavioral science more broadly:

A contextual behavioral science (CBS) approach focuses on developing adequate measures of the key processes thought to be involved in psychological difficulty and change and examines their relations to psychopathology and behavior. There need to be tight links between theoretical constructs and the auxiliaries and conditions of measurement, so that empirical problems can be attributed to the theory rather than to characteristics of the measure (Hayes et al., 2013, p. 190).

The focus on theory-driven process variables and the accurate measurement and evaluation of these variables as mediators of ACT is highly commendable. Multiple issues, however, continue to challenge the implementation of this approach. In the first half of this paper, we identify and examine five of the most important challenges, including: (a) distinguishing ACT processes in measurement contexts; (b) developing and rigorously validating measures of ACT processes; (c) the wide use of psychometrically weaker ACT process measures and the more limited use of stronger measures in past work, and the uneven study of processes; (d) the inconsistency of evidence that ACT processes are sensitive or specific to ACT or mediate ACT outcomes specifically; and (e) improving statistical power and transparency. In the second half of this paper, we offer detailed recommendations for addressing these challenges.

Many of the challenges we identify are not unique to ACT. Analyses of classic cognitive-behavioral therapy (CBT) interventions, for example, often either fail to show that theorized mediators are specific to CBT (e.g., Niles et al., 2014) or fail to robustly support theorized mediators (e.g., Arch et al., 2012). Furthermore, major critiques (e.g., Longmore & Worrell, 2007) have proposed a data-driven argument that the cognitive components of CBT—the precise components that differentiate CBT from earlier behavioral therapies—appear to be superfluous to its efficacy. Thus, although the focus of the current paper is on how these critiques apply to ACT, we also aim to contribute to a broader conversation about how to more rigorously study the processes that underlie effective psychological interventions.

To the credit of the ACT research community, a recent task force report (Hayes et al., 2021) put forth a set of recommendations for advancing contextual science that synergizes with several recommendations put forth herein. Further, numerous studies have leveraged critiques at the existing state of measuring mediators and processes in ACT (e.g., Ong, Pierce, et al., 2019). Few, however, have attempted to synthesize the critiques into a review and offer a series of recommendations based upon them, as we aim to do at present. The current effort is distinguished by focusing specifically on mediation and process variables and by reviewing the state of extant challenges and basing the recommendations specifically

¹We use this term following Hayes and colleagues (2019), who define therapeutic processes as “the underlying change mechanisms that lead to the attainment of a desirable treatment goal.”

on these challenges. We thus aim to make a unique contribution while synergizing with extant efforts.

We have used and studied ACT extensively and believe that ACT has made tremendous contributions to intervention theory, clinical application, and research. At the same time, we are also aware that programmatic research entails risks, as does focusing on any single frame or domain of inquiry (Lilienfeld, 2017). Thus, in this paper we are taking a step back in order to contribute to the goal of continuous reevaluation and constructive critique. Given that conducting a comprehensive review of processes and mediating variables across the entire ACT literature lies beyond the scope of this paper, we rely heavily on existing reviews and meta-analyses.

Part 1: The Challenges of Researching ACT Processes and Mediators

CHALLENGE I: DISTINGUISHING ACT PROCESSES

The question of how best to distinguish process or mediating variables within ACT is fundamental to evaluating whether such variables function in accordance with theory. ACT is based on a complex model most commonly defined as involving the six overlapping Hexaflex processes listed above. In this section, we outline the challenges that have arisen as researchers have attempted to distinguish these processes in measurement contexts.

First, a major strength of the ACT Hexaflex model—its capacity to be adapted to address diverse treatment targets and problem areas—also functions as a weakness at times in that it hinders the precise definition and measurement of ACT processes, opening the door to diffuse and unfocused mediation approaches. Otherwise stated, the ACT model is so flexible and multifaceted as to risk encompassing nearly any therapeutic process or outcome (except, perhaps, directly modifying problematic cognitions; Hayes et al., 2011); this adaptability renders it challenging to identify the processes that are most central to ACT. For example, prior research has framed phenomena that are not classically part of the six-process model, such as self-compassion, as key mediators in ACT (e.g., Ong, Barney, et al., 2019), introducing ambiguity about the core processes through which ACT functions. In addition, ACT incorporates techniques drawn from a variety of traditions and theories, making it hard to define at a procedural or technical level. While the challenges associated with precisely defining and validating theorized constructs in psychology and their causal relationships to one another is not unique to ACT (e.g., Eronen & Bringmann, 2021), overcoming these challenges remains critical for understanding how ACT works.

Ambiguity in distinguishing ACT processes occurs, for example, when researchers refer to the most commonly used measures of psychological flexibility—the Acceptance and Action Questionnaire (AAQ; Hayes et al., 2004) and AAQ-II (Bond et al., 2011)—variously as measures of acceptance, psychological flexibility, psychological inflexibility, or experiential avoidance, depending on the publication. To investigate this empirically, we selected two high-quality meta-analyses from a recent review of ACT meta-analyses (Gloster et al., 2020) that focused on anxiety, stress, or depression outcomes (Bluett et al., 2014; French et al., 2017) which represent common areas of inquiry. We coded the individual studies contained therein for how each defined or described the construct(s) measured by the AAQ/AAQ-II. In

the studies within the Bluett et al. meta-analysis, the AAQ/AAQ-II was most commonly defined as a measure of “experiential avoidance” or “psychological flexibility.” Also common were definitions that combined constructs—for example, “experiential avoidance and psychological acceptance,” “psychological flexibility and acceptance,” “emotional avoidance and emotion-focused inaction,” and “avoidance and immobility.” The studies in the French et al. meta-analysis showed similarly varied definitions for the AAQ and AAQ-II. Definitionally, this inconsistency in terminology is problematic because subprocesses (e.g., acceptance, experiential avoidance) cannot be equivalent to the processes under which they are nested (e.g., psychological flexibility). From an empirical perspective, if the AAQ and its variants are used to measure psychological flexibility, they should have a six-factor structure as specified in ACT theory. In addition, previous research suggests that positive and negative versions of the same construct (e.g., health; Ryff et al., 2006) and positively and negatively worded items on self-report measures (e.g., Wang et al., 2015), are typically not equivalent, which raises concerns with referring to the AAQ and AAQ-II as measures of both acceptance and experiential avoidance.

Finally, the literature does not consistently distinguish between variables that should be defined as ACT processes and those that should be defined as ACT outcomes (such as psychopathology symptoms and behaviors). For example, for most anxiety disorders, avoidance and anxiety interference are both symptoms of the disorder (and often outcomes in clinical trials) *and* target processes in ACT. Thus, in ACT trials, using a measure of avoidance or anxiety interference as a process measure while also using an anxiety symptom measure as an outcome introduces methodological and conceptual problems.

For example, well-validated, widely-used anxiety measures, such as the Overall Anxiety Severity and Impairment Scale (OASIS) and the Generalized Anxiety Disorder–7 scale (GAD-7) contain anxiety interference items that overlap considerably (on face validity) with the AAQ-II (Bond et al., 2011). Two of the five OASIS items: “How much does anxiety or fear interfere with ... your ability to do the things you need to do at work, at school, or at home?” and “... your social life and relationships?” parallel AAQ-II items: “Worries get in the way of my success,” “Emotions cause problems in my life,” and “My thoughts and feelings do not get in the way of how I want to live my life” (reverse coded). Similarly, the GAD-7 item “How much have you been bothered by ... not being able to stop or control worrying?” overlaps with the AAQ-II item “I worry about not being able to control my worries and feelings.” While again this problem is not unique to ACT (see Eronen & Bringmann, 2021) or the AAQ/AAQ-II, clearly distinguishing between ACT processes and outcomes, and establishing predictive relationships between them lies at the heart of ACT theory. Thus, these challenges must be addressed.

CHALLENGE 2: THE DEVELOPMENT AND RIGOROUS VALIDATION OF MEASURES OF ACT PROCESSES

Given the centrality of process variables in ACT, researchers initiated efforts to develop and validate process measures relatively quickly following the publication of the foundational book on ACT (Hayes et al., 1999). The first widely used, validated, and published measure of ACT processes, the AAQ (Hayes et al., 2004), was designed to measure experiential

avoidance. Though the AAQ was initially validated using data from over 2,400 participants, the authors noted that it was “likely to be relatively insensitive when used as a process measure to assess the impact of ... ACT” (p. 573), adding that “some of the items seem too complex” and the low alpha was likely “to remain an issue” and require “a more multidimensional approach” (p. 572). In that the AAQ was the first major measure of ACT processes, these challenges make sense within the developmental trajectory of measurement development.

These issues, as well as a somewhat unstable factor structure (Bond & Bunce, 2003), motivated an effort to refine the AAQ, which resulted in the development of the AAQ-II (Bond et al., 2011). This new measure was studied in a large sample and, relative to the AAQ, it evidenced higher alpha coefficients, showed good test–retest reliability, and was better at predicting multiple outcomes both cross-sectionally and longitudinally. Since the widespread uptake of the AAQ-II, however, researchers have advanced serious challenges to its validity, with a series of studies showing that the measure loads more strongly with negative affect, distress, and neuroticism measures than with third-wave behavior therapy constructs, such as mindfulness and acceptance, and does so to a greater extent than alternative measures of psychological flexibility or experiential avoidance (Allen, 2021; Rochefort et al., 2018; Tyndall et al., 2019; Wolgast, 2014). Several authors have now concluded that the AAQ-II performs more as a measure of distress/neuroticism than acceptance (Rochefort et al., 2018; Tyndall et al., 2019; Wolgast, 2014). In addition, compared to alternative ACT process measures, the AAQ-II changes less during ACT (Benoy et al., 2019), suggesting that it is also insensitive to intervention. Further, the AAQ-II has performed suboptimally in item-response theory analyses, which have found that the same score can reflect differing degrees of psychological inflexibility, and that low scores may be uninterpretable (Ong, Pierce, et al., 2019). Finally, differences in AAQ-II scores appear to be more poorly defined in treatment-seeking and student samples than in community samples, showing that across samples, score differences are not equivalent, and thus not comparable (Ong et al., 2020). Together, these findings suggest that the AAQ-II is more of a measure of outcomes or symptoms measure than of ACT processes, is comparatively insensitive to ACT intervention (compared to alternative ACT process measures), and has psychometric weaknesses at the item level. The contextual behavioral science community has recently called for studies that use denser measurement and integrate idiographic methodologies (Hayes et al., 2021)—however, the critiques of the AAQ and AAQ-II would likely extend to, and possibly even be magnified within, the context of more intensive measurement. Context-specific versions of the measures would likely have similar shortcomings.

Thus, the field remains in vital need of rigorously validated alternative measures of acceptance or experiential avoidance that are appropriate for use in nomothetic and idiographic contexts. Most of these critiques of the AAQ-II appeared 7 or more years after its publication and widespread use. To the credit of the ACT research community and consistent with the developmental trajectory of measure development, psychometrically sounder measures of ACT processes have been developed during this period, a point we turn to next.

CHALLENGE 3: THE WIDE USE OF PSYCHOMETRICALLY WEAK ACT PROCESS MEASURES, LIMITED USE OF STRONGER MEASURES, AND UNEVEN STUDY OF PROCESSES

The early development and adoption of the AAQ soon after the publication of the seminal book on ACT likely built inertia around its pervasive use and that of its descendent: the AAQ-II. Much of the earlier ACT literature used these weaker process measures, thus undermining the reliability and validity of findings on ACT processes. Numerous reviews have found that the AAQ and AAQ-II are the most frequently used process measures in ACT intervention studies, particularly in older studies. For example, a meta-analysis of the use of ACT self-help for anxiety, depression, and psychological flexibility found that in the focal studies, the AAQ/AAQ-II was the most common measure of psychological flexibility (French et al., 2017). In addition, a meta-analysis of the relationship between psychological inflexibility and anxiety (Bluett et al., 2014) focused exclusively on the AAQ/AAQ-II as “the standard measure used to assess psychological flexibility/inflexibility” (p. 613), suggesting the ubiquity of its use at that time.

If the AAQ/AAQ-II is assumed to be a measure of acceptance or experiential avoidance (see the critique of definitional issues above), then several alternative measures show superior properties across most studies. For example, the Multidimensional Experiential Avoidance Questionnaire (MEAQ; Gámez et al., 2011), as well as its briefer version, demonstrate better convergent and discriminant validity than the AAQ-II (Allen, 2021; Rochefort et al., 2018; Tyndall et al., 2019). Other work has illustrated the superior treatment sensitivity of alternative measures of psychological flexibility—namely, the AAQ-II-R, the Open and Engagement State Questionnaire, and Psyflex, in clinical samples (Benoy et al., 2019). However, these alternative measures remain less commonly used in research on ACT, likely because most of the critiques of the AAQ-II were late in coming.

The measurement of values and committed action has been similarly challenged. A systematic review of values measures (Reilly et al., 2019) identified the Valuing Questionnaire (Smout et al., 2014) and the Engaged Living Scale (Trompetter et al., 2013) as having the strongest psychometric properties (among non-pain-focused measures). However, these measures were used in only 9 and 2 studies, respectively, whereas psychometrically weaker measures, such as the Valued Living Questionnaire (VLQ; Wilson et al., 2010) and the Bull’s-Eye Values Scale (BEVS; Lundgren et al., 2012), were used in 27 and 5 studies, respectively, likely due to their earlier development. Thus, two of the three most commonly used values measures were found to be relatively weaker measures. A second systematic review of values measures (Barrett et al., 2019) also found that the Valuing Questionnaire and Engaged Living Scale were two of the most psychometrically sound values measures, although in this case, the authors also concluded that the VLQ was a strong measure despite presenting evidence of an unstable factor structure in cross-cultural samples. In addition to these psychometric limitations, many of the measures examined in these reviews do not distinguish between values and committed action processes, treating them as unidimensional. For example, research on the VLQ relies on a total score consisting of the multiplication of items assessing each value’s importance and consistency—items that presumably relate to values and committed action, respectively. Thus, the ability of the

most commonly used measures to distinguish between theoretically separable processes is unclear.

In addition, a recent review of ACT mediators (Stockton et al., 2019) concluded that the extant studies have been “overly focused on a small number of putative processes” (p. 333). For example, relative to experiential avoidance and psychological flexibility, self-as-context has been largely neglected as an object of study (Fishbein et al., 2022) and values and committed action have been comparatively understudied. This narrow focus in prior research has limited the ability to detect which processes undergo the most change following a specific ACT intervention, for whom, and in what ways, and how distinct ACT processes are related to one another and to outcomes over time.

The widespread use of psychometrically weaker measures presents a significant barrier to gaining a nuanced understanding of ACT’s processes of change and to interpreting prior research. The conceptual ambiguity surrounding measures such as the AAQ and AAQ-II obscures the processes of change that occur during ACT by poorly or inaccurately representing pathways of interest. The uptake of psychometrically strong and conceptually unambiguous measures would generate more compelling evidence for the ways in which ACT processes do, or do not, translate into therapeutic processes of change. Fortunately, as the developmental trajectory of ACT process research presses forward and many psychometrically stronger measures are available, researchers have begun shifting toward using stronger measures. Thus, the challenges outlined herein are becoming time limited as research in this area continues to advance.

CHALLENGE 4: INCONSISTENCY OF EVIDENCE FOR TREATMENT SENSITIVITY, TREATMENT SPECIFICITY, AND MEDIATION BY ACT PROCESSES

Components of Statistical Mediation—Researchers often examine whether therapy processes account for change in outcomes in predicted ways by assessing statistical mediation (e.g., MacKinnon et al., 2007).² Statistical mediation typically that must demonstrate both a unidirectional effect of the treatment condition on a putative mediational variable (the *a* path) and a unidirectional effect of the mediational variable on some outcome variable (the *b* path; Judd & Kenny, 1981). This section focuses on the effect of ACT on mediational/process variables, and thus reviews issues related to the *a* path; we direct readers to existing reviews of process–outcome associations for issues regarding the *b* path.

Two conditions are needed to obtain a significant *a* path in a randomized controlled trial (RCT) that would indicate a variable mediates ACT’s effects on outcome variables. First, the mediator variable must be *treatment sensitive*: It must change as a result of exposure to ACT. Second, the variable must also be *treatment specific* to ACT: It must change more as a result of exposure to ACT than to comparator conditions. This section outlines findings from

²The mediation framework described here has numerous drawbacks that could be overcome with idiographic analytic approaches using dense longitudinal data (Hofmann et al., 2020). However, these approaches are still in their infancy (as acknowledged by the CBS Task Force, Hayes et al., 2021). Because research on ACT has used the traditional statistical mediation approach for decades and these potential improvements have yet to be widely adopted, we focus on evidence from traditional statistical mediation analyses here. Treatment sensitivity and specificity of process measures will likely be critical for the recommended newer approaches to evaluating the mechanisms of ACT as well.

recent reviews and specific studies that shed light on the treatment sensitivity and specificity of ACT processes.

Evidence for Treatment Sensitivity—One concerning pattern is that some early widely used ACT process measures have not been evaluated or optimized as part of their development for treatment sensitivity (the first requirement for a significant *a* mediation path). For example, although numerous context- and population-specific variants of the AAQ-II have been developed, most were not tested for treatment sensitivity with ACT during their initial development (see Ong, Lee, et al., 2019, for a review). As noted above, the AAQ-II has been shown to have suboptimal treatment sensitivity (Benoy et al., 2019). In addition, a systematic review of the values questionnaire literature (Reilly et al., 2019) found that while many values measures have been assessed for treatment sensitivity after their development, widely used measures such as the BEVS and VLQ did not consistently change following ACT interventions. These two shortcomings—the lack of evaluation of treatment sensitivity in many cases, and the failure to consistently demonstrate treatment sensitivity among widely used ACT process measures—pose a major challenge to progress in the study of how ACT works. Fortunately, as psychometrically stronger measures become more widely adopted, these challenges may shift. To understand how best to move forward, however, they remain important to delineate.

Evidence for Treatment Specificity—The research on treatment specificity also reveals potential problems—namely, the evidence for the treatment specificity of psychological flexibility in randomized trials of ACT, usually operationalized by the AAQ or AAQ-II, appears to depend at least in part on ACT's comparator condition. Given the large number of studies on this topic, we focus on evidence from meta-analyses. In a meta-analysis comparing the efficacy of ACT and CBT in the treatment of anxiety and obsessive-compulsive disorders, Bluett and colleagues (2014) found that ACT did *not* lead to greater improvement on the AAQ/AAQ-II than CBT conditions. By contrast, meta-analyses focused on other comparisons—ACT versus inactive, usual care, or psychoeducation controls for family caregivers (Han et al., 2020), online ACT versus mixed inactive and active comparators for diverse clinical populations (Thompson et al., 2021), and self-help ACT versus mixed inactive and active comparators for diverse clinical and nonclinical populations (French et al., 2017)—showed greater improvement overall in ACT conditions on the AAQ/AAQ-II and related measures. However, the evidence is varied, as a small meta-analysis of ACT interventions for reducing burnout among direct-care staff found no improvement on the AAQ/AAQ-II among studies comparing ACT to psychoeducational or wait-list control conditions (Reeve et al., 2018). In brief, while research has shown that ACT sometimes increases scores on the AAQ/AAQ-II more than inactive or non-gold-standard controls, the evidence comparing ACT to gold-standard psychotherapy comparators is more mixed. These results may be due to the AAQ/AAQ-II functioning primarily as a measure of distress rather than psychological flexibility (Rochefort et al., 2018; Tyndall et al., 2019; Wolgast, 2014). More psychometrically rigorous measures may show better treatment specificity—however, this is difficult to determine from current studies.

Compared to evaluations of treatment specificity for psychological flexibility, evaluations of treatment specificity for the six ACT processes are less common. One meta-analysis (Han et al., 2020) that focused on ACT interventions for family caregivers examined treatment specificity for cognitive defusion, values consistency, and present-moment awareness. The findings for treatment specificity were largely null—however, they analyzed these processes across only a very small number of studies (two to four per process). The systematic review by Stockton and colleagues (2019) identified just four studies examining cognitive defusion as a mediator of ACT, including three that compared ACT to CBT or cognitive therapy in anxious or depressed samples, which together showed mixed findings. Notably, the included studies used a variety of measures to assess defusion, complicating attempts to synthesize the findings. They identified one small study examining mediation with values consistency (Lundgren et al., 2008) that showed treatment specificity of values consistency in ACT relative to supportive psychotherapy—however, a recent, larger RCT of ACT versus minimally enhanced usual care for anxious cancer survivors did not observe treatment specificity for values consistency (Fishbein et al., 2022). Stockton and colleagues identified only one study examining mediation with present-moment awareness that supported treatment specificity for ACT compared to cognitive therapy (Forman et al., 2007), and no studies examining mediation with self-as-context. In a subsequent systematic review of research on self-as-context, Godbee and Kangas (2020) identified only one very small controlled (but nonrandomized) trial of a multisession intervention that evaluated self-as-context, concluding that such research is “in its infancy” (p. 930). For committed action, Stockton and colleagues identified just two studies that evaluated committed action as a mediator of ACT, neither of which supported treatment specificity.

In summary, particularly when ACT is compared to CBT, the evidence that ACT processes are specific to ACT is mixed, and studies that evaluate the specificity of ACT processes beyond psychology flexibility are sparse. These findings align with the possibility that ACT’s mechanisms are shared with those of other efficacious psychological interventions (Arch & Craske, 2008; Gaudiano, 2011). These findings also converge with a recent meta-analysis of mediators of acceptance and mindfulness-based therapies more broadly (Johannsen et al., 2022) that concluded that “mediator specificity could not be established,” though more research is needed. Moreover, the findings should be interpreted with some caveats.

Caveats and Limitations—We note four caveats in interpreting the findings on treatment sensitivity and specificity. First, the response to ACT intervention content may differ across distinct populations and delivery formats (e.g., group vs. individual, app vs. in person) and thus participants may experience different degrees and patterns of change across ACT process variables. Second, clinicians vary in their training in and fidelity to the ACT model, which likely affects the strength of the intervention’s impact on ACT process variables. Third, different measures of the same ACT process may capture treatment effects to greater or lesser extents (e.g., Benoy et al., 2019). Finally, emerging evidence from daily diary studies indicates that ACT processes can vary considerably over a span of days or weeks (Pavlicic et al., 2021)—thus, more global process measures that are administered at intervals of multiple months likely do not fully capture change over time. This is a problem

with self-report measures, especially trait measures, within contextual intervention contexts (Newsome et al., 2019). However, in that the vast majority of ACT clinical trials have relied largely or exclusively on self-report measures of ACT processes, such measures constitute most of our knowledge of ACT mediation to date.

CHALLENGE 5: PROBLEMS WITH STATISTICAL POWER AND TRANSPARENCY

The replication crisis in psychology (Open Science Collaboration, 2015) has highlighted methodological and analytic practices that prevent researchers from obtaining reliable results across similar studies. Such practices are common in clinical science as well (Tackett & Miller, 2019) and are also likely present in a portion of ACT studies, although little research has documented meta-science concerns directly in this literature. Nevertheless, small and unfunded clinical trials and correlation studies that comprise much of the ACT literature are particularly vulnerable. Further, while large and externally funded ACT clinical trials often preregister their primary analyses and sometimes publish the trial protocol prior to the study, many do not preregister process or mediational analyses. Furthermore, broader evidence from a recent review of mediational analyses across five leading psychology journals and diverse behavioral interventions concluded that “the likelihood of [questionable research practices] in tests of mediation was high—perhaps even a foregone conclusion” (Götz et al., 2021, p. 106). In sum, it seems likely that these broader problems in scientific practice affect some unknown portion of ACT studies on mediation. There is a clear need for meta-scientific studies of ACT research, specifically in relation to mediation (i.e., Götz et al., 2021), to study the extent of these problems in this literature.

Examples of questionable research practices (QRPs) in the larger literature are numerous, and these practices are more likely to be used in studies that are not preregistered or based on prepublished protocols (Tackett & Miller, 2019). Harmful practices include the selective reporting of conditions, measures, studies, participants, data, and other information, which makes it difficult to fully evaluate a study. In addition, researchers often engage in analytical practices that capitalize on chance to obtain positive findings, such as *p* hacking—repeatedly running statistical analyses and only reporting those that are significant—and “the garden of forking paths”—running many analyses but only reporting a subset (Friede & Frankenbach, 2020; John et al., 2012). Similarly, researchers often do not distinguish between exploratory and confirmatory analyses, making it difficult to understand the likely rate of false positives in studies (which are expected to be higher in exploratory research). One consequence of this approach is that unexpected results are often reported as if they were hypothesized a priori (Kerr, 1998). These practices, as well as file drawer problems (not publishing studies with null results), increase false positives in the clinical literature and inflate parameter estimates of the associations between variables. For example, a large-scale replication of experimental psychology studies found that while 97% of results in the initial studies were statistically significant, only 36% of results in the replication were significant, even when larger samples were used (Open Science Collaboration, 2015), and at least one prominent website that tracks the publication of clinical trials suggests that 25% of clinical trials do not report their findings (<https://fdaaa.trialstracker.net>).

Because a large portion of articles on ACT mediational analyses use data collected prior to the “replicability revolution” or were not preregistered, it is likely that many suffer from one or more of these problems. In addition, statistical power seems to be a particular concern. In the Stockton et al. (2019) meta-analysis that screened for some of the highest-quality studies of mediation of ACT processes, the average total sample size was $n = 70$, with over half of the studies having fewer participants. Given that these were selected as higher-quality studies, it is likely that other mediational analyses have even smaller sample sizes on average. A simulation of the relationship between sample size and power in mediational analyses conducted by Fritz and MacKinnon (2007) indicated that small effects—which are common in the mediation literature—typically required samples of several hundred participants to have adequate power, suggesting that many published mediational studies are underpowered. In that many ACT trials to date have involved modest samples (Hayes, 2022), inadequate sample size may particularly affect mediational analyses based on these trials. In sum, the combination of low statistical power and pervasive problematic research practices in the psychological literature points toward the possibility that many of the results reported in the ACT mediational literature (and other psychotherapy literatures) are unreliable and will not replicate due to meta-scientific problems.

Part 2: Recommendations in Research on ACT Processes and Mediators

Table 1 summarizes the challenges and recommendations for research on ACT processes and mediators.

RECOMMENDATIONS FOR CHALLENGE I: DISTINGUISHING ACT PROCESSES

Empirically distinguishing the structure and relationships of ACT process variables remains a foundational step in measuring them. ACT has been defined as having six core Hexaflex processes that are central to psychological flexibility (Hayes et al., 1999, 2012). This model is widely used in teaching (Harris, 2019) and studying ACT (Hayes et al., 2013), and Hayes and colleagues (1999, 2012) provide conceptual definitions of each process. Validating the Hexaflex model thus provides an excellent starting point and foundation for understanding ACT. Yet surprisingly little work has empirically evaluated the structure of the Hexaflex model. If its six lower-order processes are nested under the higher-order construct of psychological flexibility, as theorized, then their measurement should reflect this structure—or should challenge and revise it.

First, a measure of psychological flexibility with subscales for each lower-order Hexaflex process could be developed. The Psy-Flex (Gloster et al., 2021), for example, measures all six Hexaflex processes with one item each and demonstrates a one-factor structure—a strength for usability and simplicity, and a helpful evaluation of the Hexaflex structure. However, the Psy-Flex would have difficulty distinguishing the unique contribution of distinct processes (due to using only one item per process). Thus, alternatively, researchers could select (or develop, as needed) sound measures for each lower-order process, including behaviorally based measures. Then the Hexaflex structure could be formerly evaluated in diverse populations using factor analyses with multi-item measures for each process (see Flora & Flake, 2017; Sellbom & Tellegen, 2019, for relevant guidance on this topic). If

the Hexaflex factor structure is not consistently supported as theorized (i.e., six specific constructs that are facets of one higher-order construct), it could be updated and revised. For example, given their conceptual overlap, one could evaluate whether the Hexaflex processes of self-as-context and present-moment awareness are empirically distinct. If self-as-context and present-moment awareness items prove difficult to distinguish at the level of measurement, but load onto the same higher-order latent construct of psychological flexibility, it would provide empirical motivation to collapse those two processes into one and thus simplify ACT's model. Alternatively, the original Hexaflex could prove to serve primarily as a clinical model for teaching and understanding ACT, but a different model would emerge for how ACT's processes function at the empirical level. In addition, measurement invariance analyses (see Millsap, 2012; Putnick & Bornstein, 2016) could help the CBS field to evaluate whether constructs, and the scales that measure them, can be applied equally across populations, versus whether different models and scales are needed for different therapeutic contexts or sociocultural contexts.

Second, we must consider the issue of process specificity. Change in Hexaflex processes should predict change in therapeutic outcomes. ACT interventions also are likely to change processes implicated in many forms of psychotherapy, such as a sense of common humanity or therapeutic alliance—however, the ACT model would predict that ACT-specific processes, rather than these broader processes, should be the key to improving ACT outcomes. In addition, we would expect ACT processes to improve more and to better predict outcomes following ACT than processes from other cognitive-behavioral models, such as change in dysfunctional beliefs from CBT. Evaluating this will require routinely assessing multiple process variables in ACT trials (including both ACT-specific processes and broader processes or those from other treatment models) and distinguishing them as such, to clarify the degree to which ACT-theorized processes are specific to ACT or overlap with those of other therapies (see Arch & Craske, 2008). While some studies have already evaluated process variables in this manner, for example, in the anxiety disorder and related symptom literature (Arch et al., 2012; Forman et al., 2007), more data are needed to draw strong conclusions.

Third, ACT researchers should also be clearer about specifying processes versus outcomes in mediational models and should explain how ACT theory supports the proposed relationships. For example, if ACT improves cognitive defusion (the process), does ACT theory predict that reductions in suffering (the outcome) should occur in the same time period or later? Also, if an outcome is an overt behavior (e.g., taking medication), how does the behavior relate to psychological flexibility (the process), especially when, as in the case of medication adherence and many other overt behavioral outcomes, the behavior can be understood as a committed action and thus as a lower-order process of psychological flexibility? Our review also suggests that the conceptual distinctions between internal shifts in ACT processes, such as acceptance and external shifts in overt behavior, often have not been made clearly and consistently in the literature (see though Villatte et al., 2016, for a notable exception). For example, relying on a single unidimensional process measure of psychological flexibility measures it in a manner that merges internal shifts in ACT process and external shifts in overt valued behavior. This renders it difficult to discern which ACT processes are most active in a given intervention, or the extent to which increased

commitment to pursue valued actions (a process) overlap with the behavioral outcome of interest (e.g., taking one's medication). Drawing clearer conceptual lines between constructs reflecting psychological processes versus outcomes, and among the lower-order processes, has a strong conceptual start in ACT, but may at times require a return to more formative work in theoretical construct definition (Clark & Watson, 2016, 2019) before embarking on the methodological and statistical approaches suggested in the sections that follow.

The recent task force on CBS research strategies (Hayes et al., 2021) commendably argues for more idiographic, contextually based approaches to measuring process variables in contextual behavioral therapies, such as ACT. We hope that ACT researchers build on this report by explicating practical strategies for accomplishing the goals proposed by the task force as well as those outlined herein, including conducting studies that measure multiple processes (both specific and not specific to ACT) and clearly defining and differentiating between the different types of examined processes and between processes and outcome in theoretically rigorous ways.

RECOMMENDATIONS FOR CHALLENGE 2: THE DEVELOPMENT AND RIGOROUS VALIDATION OF MEASURES OF ACT PROCESSES

Conclusions about process variables in psychotherapy are only as valid and reliable as the measures themselves. As noted above, independent research groups have seriously challenged the validity and reliability of ACT's most widely used process measures: the AAQ and AAQ-II (Allen, 2021; Ong et al., 2020; Rochefort et al., 2018; Tyndall et al., 2019; Wolgast, 2014). What guidelines can be applied to more rigorously develop and validate ACT process measures?

First, on a conceptual level, we recommend that ACT process measures more precisely specify what is being assessed and assess only one process per scale (or per subscale). ACT process measures have been moving in this direction—for example, the MEAQ (Gámez et al., 2011) and the Cognitive Fusion Questionnaire (Gillanders et al., 2014) each measure only one process in the ACT model (though the former does so multidimensionally). Second, both conceptually and psychometrically, ACT process measures must be clearly distinguishable from outcome and symptom measures. For example, if the ACT process measure is “experiential avoidance” and the outcome is “anxiety symptoms,” then experiential avoidance-related items should not conceptually overlap with the anxiety measure. Alternatively, researchers should consider using an anxiety symptom measure that focuses on other relevant dimensions of anxiety (e.g., worry, sympathetic arousal). At the most basic level, this determination could be based on the face validity of the items (e.g., which ones appear to overlap in content) or at a more sophisticated level, on conducting a factor analysis and checking that process scale items do not cross-load onto a latent factor capturing the outcome variable construct or vice versa (e.g., Flora & Flake, 2017). Third, process variables should be responsive to intervention and should statistically mediate outcomes in analyses using either nomothetic mediation approaches or idiographic approaches recommended by Hayes and colleagues (2021). Fourth, an ACT process measure should ideally be validated in the population of interest and be responsive to ACT intervention within that population. For example, if a measure was developed and

evaluated only in online samples or among undergraduates, and researchers then wish to use it in an intervention trial of adults with depression, it would be recommended to first examine the psychometric properties and factor structure of that scale in the new population (i.e., engage in “ongoing validation” per Flake et al., 2017). Such pilot work could critically demonstrate, before engaging in a lengthy and expensive clinical trial, whether the scale reflects the construct of interest, and whether comparisons of scores across populations are warranted.

Fifth, as part of a contextually based approach, ACT process measures should ideally be assessed within the temporal contexts that are most relevant for that population—for example, in the aftermath of a panic attack for those with panic disorder. Thus, in addition to conducting ongoing work at the trait level, researchers should also measure ACT processes as momentary states in relevant contexts (see Hayes et al., 2021). Kashdan et al. (2014) provided an excellent example of using a contextual approach to assess experiential avoidance. Finally, measures would benefit from being tailored to reflect the specific conceptual issues involved in measuring each ACT process, as Barney et al. (2019) recently provided for the measurement of the “valuing process” in ACT.

On a psychometric level, published guidelines provide clear recommendations for the development and evaluation of self-report measures—the most prominent example is the consensus-based standards for the selection of health measurement instruments (COSMIN) guidelines (Mokkink et al., 2010), including recent COSMIN guidelines developed for evaluating content validity (Terwee et al., 2018). Researchers should also evaluate whether their measures are invariant across study groups of interest and longitudinally over time, including pre- to postintervention (Millsap, 2012). COSMIN guidelines set a very high bar and thus an ACT process measure does not necessarily need to meet every guideline to be considered valid (we suspect that no current ACT measures and few other measures would; see Serowik et al., 2018). Further, COSMIN does not cover every area of interest to ACT researchers. Nonetheless, these guidelines set an important and respected standard. The extent to which a measure’s development and validation process follows these guidelines can help researchers and clinicians identify the most rigorously validated measures. Given the complexity of assessing psychometric quality, we recommend collaborating with a statistical expert. Meeting as many COSMIN guidelines as makes sense for a given set of validation goals would raise the quality of ACT process measures.

In addition, as noted by the CBS Task Force (Recommendation 19; Hayes et al., 2021), self-report instruments currently in use for measuring ACT processes are not necessarily suitable for intensive longitudinal study, such as in experience sampling designs. Researchers need to evaluate the psychometric properties of current ACT process measures when employed in intensive longitudinal data collection contexts, to develop and validate new measures as needed, and to stay abreast of developments in psychometric analysis for measurement in those contexts.

Another important next step is to develop more objective instruments to measure ACT processes. Some have argued that the overreliance on selfreport measures, as opposed to measures of overt behavior, is discordant with the mission of ACT and of contextual

behavioral science broadly (Newsome et al., 2019). Thus, developing more objective and behaviorally based instruments of ACT processes remains critical to ACT's scientific aims (Hayes et al., 2021). Such instruments could reflect both laboratory-based behavioral tasks and more ecologically valid measures of behavior performance. Researchers have developed laboratory-based instruments to measure aspects of mindfulness (see Hadash & Bernstein, 2019), and these tasks could perhaps be used or altered to measure some ACT processes. For example, a task developed by Hadash et al. (2016) may provide a starting point for a laboratory-based self-as-context instrument. Likewise, researchers can employ in-the-moment assessment of behaviors, such as experience sampling of whether someone is being mindful at the moment of assessment, as well as passively collected data reflecting potentially values-related behaviors, such as exercising, visiting locations associated with valued activities, or taking medication. The simultaneous modeling of self-report data collected at critical study time points with laboratory or ecological assessments can improve overall measurement accuracy and determine whether and how self-reporting on ACT processes differs from behavioral observations (e.g., Schloss & Haaga, 2011).

RECOMMENDATIONS FOR CHALLENGE 3: THE WIDE USE OF PSYCHOMETRICALLY WEAK ACT PROCESS MEASURES, LIMITED USE OF STRONGER MEASURES, AND UNEVEN STUDY OF PROCESSES

Across most contexts, the first step in improving the measurement of ACT processes is to shift away from using measures whose psychometric properties have drawn serious criticism, including the AAQ and AAQ-II, or have been identified as weaker than multiple alternatives in a given domain, such as, according to one review (Reilly et al., 2019), the BEVS. Although these measures were commendable initial efforts, researchers must now move toward more psychometrically sound, treatment-sensitive, clearly interpretable, and precisely targeted process measures. As improved measures have become increasingly available, this effort is well underway. Researchers should consider which ACT process(es) are most likely to change following the focal intervention and measure those processes specifically, rather than measure psychological flexibility as a whole by default. In addition, researchers and clinicians should be nimble—adopting new measures quickly when they are shown to be psychometrically or conceptually superior to previous ones—to avoid the inertia now evident in the persistent use of the same suboptimal measures.

Recent efforts to systematically evaluate the quality of ACT process measures using COSMIN or other criteria (e.g., Reilly et al., 2019) are an important step toward identifying stronger process measures and should be continued. To build on these efforts and to promote the use of the strongest process measures on a larger scale, the ACT research community should consider undertaking more coordinated efforts. Research organizations, such as the Palliative Care Research Cooperative Group (PCRC; Abernethy et al., 2010), have advanced the use of high-quality measures by developing and disseminating an instrument library of relevant robust measures in their field. The PCRC library evaluates each measure along numerous dimensions, including measure quality, and provides relevant citations. In addition, the main clinical research funder in the United States, the National Institutes of Health (NIH), led an extensive collaborative effort to provide high-quality standardized

measures across numerous outcomes relevant to mental and physical health (Cella et al., 2007).

These broader approaches provide important models for how to improve measurement on a large scale. Although psychotherapy communities tend to lack the funds necessary to undertake an effort on the scale of the NIH endeavor, ACT researchers, via the Association for Contextual Behavioral Science (ACBS), have long modeled freely sharing measures through the association's website and listserv. The ACT community within ACBS should consider adopting the approach taken by the PCRC's instrument library, which would facilitate a shift toward the routine adoption of more psychometrically sound ACT process measures in research, grant writing, clinical, and other applied settings. In addition, research teams or ACBS could support efforts to identify the best process measures for a given context, and then encourage researchers to use these measures across multiple independent studies (in similar populations using the same preregistered statistical approaches) to facilitate a more systematic approach to building knowledge about ACT processes. Overall, we encourage ACT researchers to devote as much effort to selecting high-quality process measures as to designing their studies. There are encouraging signs that many are moving in that direction, and we applaud them. After all, the resulting conclusions will only be as strong and credible as the measures themselves.

RECOMMENDATIONS FOR CHALLENGE 4: THE INCONSISTENCY OF EVIDENCE FOR TREATMENT SENSITIVITY, TREATMENT SPECIFICITY, AND MEDIATION BY ACT PROCESSES

In addition to developing and selecting high-quality measures for an ACT process in a given research context, researchers should also consider two specific aspects of their study design to more accurately evaluate the treatment specificity and sensitivity of ACT processes. Our first recommendation is that researchers should consider the time scale over which they expect ACT processes to change in response to intervention, as well as the time scale over which they expect the focal processes to influence and be influenced by outcome variables. Mediation should assess the relationship between variables over time, rather than at a single time point (e.g., Jose, 2016). Researchers often account for time in mediation analyses of RCT data by modeling assessments of process and outcome variables from multiple time points that are weeks or months apart. Yet recent research has demonstrated that, when measured via daily diary methods, both acceptance (Pavlicic et al., 2021) and values consistency (Finkelstein-Fox et al., 2020) vary considerably more on a daily basis than may have been expected. These findings imply that if these variables are assessed only once every few weeks or months, then potentially meaningful changes in ACT process variables are likely to be lost. Moreover, as noted above, if the timing of assessments is too coarse to capture the lagged effects of change in processes upon change in outcomes, then it will be difficult to establish the temporal order (and thus, causal sequencing) of ACT's effects. This is similarly the case if assessing the bidirectional influences over time of processes and outcomes.

No methodological or statistical approach will be a panacea for resolving these challenging issues, but many could help. The timing of causal effects from one variable to another could

be explored using a combination of intensive longitudinal data collection and “idiographic” analytic modeling approaches, as explicated in CBS Task Force Report Recommendation 19 (Hayes et al., 2021). This program of research is thoroughly outlined by Hayes and colleagues (2019). Further, CBS researchers may consider applying models that capture longitudinal change on two or more variables simultaneously across a smaller number of time points (e.g., more than three, but fewer than in intensive longitudinal studies), with lagged or concurrent change, to examine how change unfolds across processes and outcomes over time (see Goldsmith et al., 2018; Usami et al., 2019, for reviews). In addition to requiring fewer time points, an advantage of these latter models is that they can be applied in contexts in which it is not practical to adopt intensive longitudinal data collection, while still offering insight into the timing and direction of process–outcome relationships.

We caution, however, that these newer models (as well as classic mediation models) may be challenging to adequately power (Fritz & MacKinnon, 2007) and fit (see McArdle & Grimm, 2010, for a tutorial) in clinical studies and trials. Specific challenges relevant (but not unique) to typical ACT research contexts range from recruiting sufficient participants to adequately powering the analysis to challenges that emerge as a consequence of psychological intervention (Fried et al., 2016), including participant changes in understanding or attitude toward a construct, and changes in variability in scores on a construct. Further, researchers must measure and model all reasonably relevant variables, or risk obfuscating the causal relationships present within the data (Rohrer et al., 2021).

Our second recommendation in improving assessment of mediation in ACT is to consider whether and how intervention fidelity to ACT is achieved in the study context. The degree of ACT fidelity in a given trial likely impacts participants’ improvement on ACT processes. For trials using human interventionists, it is critical to establish not only whether the interventionist covered certain content in a given session, but also *how* they covered it, and whether they consistently modeled the components of psychological flexibility in doing so (Plumb & Vilardaga, 2010). The ACT Fidelity Measure (O’Neill et al., 2019) is a recently published, validated instrument that can help capture this quality of ACT interventions. In addition, studies of interventionists’ fidelity to motivational interviewing interventions indicate that occasional supervision improves fidelity beyond initial training (Schwalbe et al., 2014); the same could be true of ACT. Thus, further research is needed to establish optimal ACT training and supervision schedules within a given context.

RECOMMENDATIONS FOR CHALLENGE 5: PROBLEMS WITH STATISTICAL POWER AND TRANSPARENCY

A variety of commonly recommended practices for increasing replicability, transparency, and openness would likely improve research on mediation related to ACT processes. Most of these practices involve researchers thinking beyond the results of their own study to consider how their research practices fit into a larger field of science aimed at creating reliable principles that have precision, scope, and depth. Unfortunately, many common nomothetic research practices produce statistical parameter estimates with extreme variability and poor characterization of the reliability of those estimates. While CBS researchers have argued for a more idiographic approach to research that recognizes the limitations of generalizing

from group results to individuals (Hayes et al., 2019), and while novel statistical methods are being developed that will enable such an approach (Piccirillo & Rodebaugh, 2019), the field nevertheless lacks widely used and accepted methods that meet these aims, and researchers continue to use group analyses and associated statistical methods. Even if workable idiographic methods are developed, they will benefit from most of the practices outlined below for increasing transparency and openness. We suggest solutions that will not only increase the validity and generalizability of individual studies but also increase the ability to draw conclusions across the broader literature on ACT processes of change by reducing the noise and increasing the signal. Due to space limitations, we are able to provide only a brief overview of the possible meta-science strategies that ACT researchers could adopt to improve the quality of the science on this topic. We refer researchers interested in a more general introduction to these practices in relation to clinical science to a special section in *Journal of Abnormal Psychology* (Tackett & Miller, 2019) and those who are specifically interested in methods to improve mediational studies to the Götz et al. (2021) review, which is a key article that any ACT researcher considering conducting mediational analyses would benefit from reading. In addition, a variety of other papers recommend means to improve mediational models, including how to calculate effect sizes (Lachowicz et al., 2018; Mio evi et al., 2018) that can facilitate a move away from a focus on null effect significance testing (Götz et al., 2021) and recommendations for best practices (e.g., Fairchild & McDaniel, 2017; Jose, 2019).

A primary way to reduce researcher “degrees of freedom” that can inadvertently generate bias is the use of preregistration. In preregistration, researchers publicly declare their study methods and analyses on open science sites, such as osf.io/prereg or clinicaltrials.gov (for clinical trials). Preregistration, while not perfect, does appear to increase transparency in reporting, help researchers consider appropriate sample sizes, and generally improves study planning (Toth et al., 2021). Preregistration practices occur on a spectrum (Benning et al., 2019), ranging from preregistering a study before data collection ever begins (as often occurs on clinicaltrials.gov) to co-registration (registering analyses during data collection) to postregistration (registering specific analyses after the data have been collected but before the data are analyzed). Each of these methods has particular advantages, but all serve to restrict researcher degrees of freedom in ways that will reduce the risk of capitalizing on chance in a manner that results in misleading parameter estimates. In addition, preregistration can vary from publishing very specific guidelines outlining what will happen to making relatively cursory declarations that do little to restrict degrees of freedom.

At this stage in the research on mediation in ACT, we encourage researchers to move from exploratory analyses to well-powered, preregistered confirmatory studies in which they predict the results they expect to find in mediational models based on previous findings and theory. A key goal of CBS is prediction, and preregistered predictions are a powerful means of testing predictive ability. We recommend that researchers estimate sample size requirements for mediational models prior to data collection (Schoemann et al., 2017), which addresses many but not all problems with mediation (see Götz et al., 2021). Further, rather than analyzing 10 different mediators and reporting the one significant result, researchers should evaluate only mediators that are most conceptually relevant to the focal

intervention and have been most strongly supported in prior studies using strong measures of relevant mediators. As noted in Götz et al., measurement error is particular problematic in mediation analyses and thus using assessments with greater validity and reliability should reduce measurement error (noise) and thereby increase statistical power. To evaluate mediator specificity, researchers can combine multiple mediators in multiple mediator models that include putative mediators specified by the ACT model as well as those that are broader or unspecified by the ACT model, or by using latent variables. If an analysis was not preregistered, the researcher should be explicit about the steps of the research process: Describe all measures collected, all analyses conducted (even if doing so requires an online appendix), all conditions used, and all methods for handling data missingness and outliers. Researchers should also note whether the analyses were confirmatory or exploratory, because the former type of research generally produces parameter estimates that are less affected by capitalization on chance.

In the interest of reducing file drawer effects and obtaining more reliable parameter estimates, we also recommend that researchers and journals publish nonsignificant findings, particularly for well-constructed and well-powered studies. If findings cannot be published in peer-reviewed journals due to bias toward publishing only statistically significant results (Simonsohn et al., 2014), researchers should consider filing a report on sites such as PsyArXiv so that the results can be included in future reviews or analyses. The publication of nonsignificant findings is important to identify boundary conditions for established effects and to inform future research (Landis et al., 2014) and the value of nonsignificant results should be discussed in discussion sections as a way to advance scientific practice.

Communal strategies can also be used to increase statistical power. Because clinical studies are often more expensive than other forms of research, it may be particularly important to seek communal solutions that allow for bigger sample sizes and the aggregation of data across multiple studies. For example, multisite collaborations could facilitate larger sample sizes and the use of core process measures. Open data sharing via sites such as osf.io provides another important strategy (Kryptos et al., 2019). Given the small sample sizes in most mediational analyses of ACT processes, sharing data may allow researchers to eventually combine data across studies and thus obtain more statistical power. Sharing data would be a particularly effective strategy if CBS researchers agreed on a core set of mediational measures from an instrument library that could be utilized across studies (see the previous section). We recommend sharing the analytic syntax or code that accompanies the databases and providing clear variable definitions to make it easier for others to conduct parallel analytic models or replicate and extend analyses.

As research on ACT matures, it is important to focus more on winnowing the wheat from the chaff rather than only on pushing into new research areas and new samples. With over 900 randomized trials of ACT interventions and scores of mediational analyses that have produced novel results, it is time to focus more energy on determining which of these results are highly reliable and which are not. In terms of mediation, researchers can collaborate to identify the strongest findings regarding mediation and then attempt to replicate these findings in high-powered, preregistered studies that follow mediation best practices (Fairchild & McDaniel, 2017; Götz et al., 2021; MacKinnon et al., 2007). In

addition to the methods reviewed above that focus on increasing statistical power, openness, and transparency, a number of other statistical methods and analyses are important to consider (Götz et al., 2021). Furthermore, reviewers of mediational findings should consider the statistical power of studies when interpreting findings and consider newer methods for quantitatively synthesizing them (Cheung, 2022).

As the flagship journal for ACT research, we commend the *Journal of Contextual Behavioral Science* for recently adopting a number of these open science principles (https://contextualscience.org/news/adoption_of_open_science_recommendations). Showing that the most robust findings can replicate will go a long way toward demonstrating the robustness of findings in a field with many small, underpowered, and poorly controlled studies.

Conclusion

One of ACT's central premises is that this therapeutic approach works according to its theorized processes, particularly those defined in the six-process Hexaflex model. Creating reliable and valid measures of these processes and determining the extent to which they account for positive outcomes and specifically link to ACT remain fundamental endeavors for ACT researchers. This paper considered five central challenges that limit the validity and reliability of many earlier findings on ACT processes and mediators. The recommendations we propose to meet these challenges represent very high standards for any single study or research team to meet, yet they may be necessary for the field to make progress and many ACT researchers have already begun moving in these directions. Community responses in which ACT researchers jointly create stronger standards and guidelines for investigating processes, testing mediation, and establishing collaborations that allow for greater rigor and power, provide promising paths forward. The goal is not to establish a standard that is impossible for researchers to meet but rather to provide suggestions for how CBS can grow incrementally by improving its process- and mediation-related research practices, reducing error and increasing a reliable signal in relevant studies, and facilitating the ability to combine results across studies. These goals remain vital to advancing our understanding of ACT and CBS, and they are within reach.

Acknowledgments

This work was supported by the National Institutes of Health [grant R01NR018479] to J.J.A. Author J.J.A. receives research funding from the National Institutes of Health and the National Comprehensive Cancer Network/AstraZeneca, including for research on ACT.

References

- Abernethy AP, Aziz NM, Basch E, Bull J, Cleeland CS, Currow DC, Fairclough D, Hanson L, Hauser J, Ko D, Lloyd L, Morrison RS, Otis-Green S, Pantilat S, Portenoy RK, Ritchie C, Rocker G, Wheeler JL, Zafar SY, & Kutner JS (2010). A strategy to advance the evidence base in palliative medicine: Formation of a palliative care research cooperative group. *Journal of Palliative Medicine*, 13(12), 1407–1413. [PubMed: 21105763]
- Allen MT (2021). An exploration of the relationships of experiential avoidance (as measured by the AAQ-II and MEAQ) with negative affect, perceived stress, and avoidant coping styles. *PeerJ*, 9, e11033. [PubMed: 33717714]

- Arch JJ, & Craske MG (2008). Acceptance and commitment therapy and cognitive behavioral therapy for anxiety disorders: Different treatments, similar mechanisms?. *Clinical Psychology: Science and Practice* 15(4), 263–279.
- Arch JJ, Wolitzky-Taylor KB, Eifert GH, & Craske MG (2012). Longitudinal treatment mediation of traditional cognitive behavioral therapy and acceptance and commitment therapy for anxiety disorders. *Behaviour Research and Therapy*, 50(7–8), 469–478. [PubMed: 22659156]
- Barney JL, Lillis J, Haynos AF, Forman E, & Juarascio AS (2019). Assessing the valuing process in acceptance and commitment therapy: Experts' review of the current status and recommendations for future measure development. *Journal of Contextual Behavioral Science*, 12, 225–233. [PubMed: 31828006]
- Barrett K, O'Connor M, & McHugh L (2019). A systematic review of values-based psychometric tools within acceptance and commitment therapy (ACT). *The Psychological Record*, 69(4), 457–485.
- Benning SD, Bachrach RL, Smith EA, Freeman AJ, & Wright AG (2019). The registration continuum in clinical science: A guide toward transparent practices. *Journal of Abnormal Psychology*, 128(6), 528–540. [PubMed: 31368732]
- Benoy C, Knitter B, Schumann I, Bader K, Walter M, & Gloster AT (2019). Treatment sensitivity: Its importance in the measurement of psychological flexibility. *Journal of Contextual Behavioral Science*, 13, 121–125.
- Bluett EJ, Homan KJ, Morrison KL, Levin ME, & Twohig MP (2014). Acceptance and commitment therapy for anxiety and OCD spectrum disorders: An empirical review. *Journal of Anxiety Disorders*, 28(6), 612–624. [PubMed: 25041735]
- Bond FW, & 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. [PubMed: 14640816]
- Bond FW, Hayes SC, Baer RA, Carpenter KC, Guenole N, Orcutt HK, Waltz T, & Zettle RD (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire–II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, 42(4), 676–688. [PubMed: 22035996]
- Cella D, Yount S, Rothrock N, Gershon R, Cook K, Reeve B, Ader D, Fries JF, Bruce B, & Rose M (2007). The Patient-Reported Outcomes Measurement Information System (PROMIS): Progress of an NIH Roadmap cooperative group during its first two years. *Medical Care*, 45(5, Suppl. 1), S3.
- Cheung MW (2022). Synthesizing indirect effects in mediation models with meta-analytic methods. *Alcohol and Alcoholism*, 57(1), 5–15. [PubMed: 34190317]
- Clark LA, & Watson D (2016). Constructing validity: Basic issues in objective scale development. In Kazdin AE (Ed.), *Methodological issues and strategies in clinical research* (pp. 187–203). American Psychological Association.
- Clark LA, & Watson D (2019). Constructing validity: New developments in creating objective measuring instruments. *Psychological Assessment*, 31(12), 1412–1427. [PubMed: 30896212]
- Eronen MI, & Bringmann LF (2021). The theory crisis in psychology: How to move forward. *Perspectives on Psychological Science*, 16(4), 779–788. [PubMed: 33513314]
- Fairchild AJ, & McDaniel HL (2017). Best (but oft-forgotten) practices: Mediation analysis. *American Journal of Clinical Nutrition*, 105(6), 1259–1271. [PubMed: 28446497]
- Finkelstein-Fox L, Pavlacic JM, Buchanan EM, Schulenberg SE, & Park CL (2020). Valued living in daily experience: Relations with mindfulness, meaning, psychological flexibility, and stressors. *Cognitive Therapy and Research*, 44(2), 300–310.
- Fishbein JN, Baer RA, Correll J, & Arch JJ (2022). The Questionnaire On Self-transcendence (QUEST): A measure of trait self-transcendence informed by contextual cognitive behavioral therapies. *Assessment*, 29(3), 508–526. [PubMed: 33371714]
- Flake JK, Pek J, & Hehman E (2017). Construct validation in social and personality research: Current practice and recommendations. *Social Psychological and Personality Science*, 8(4), 370–378.
- Flora DB, & Flake JK (2017). The purpose and practice of exploratory and confirmatory factor analysis in psychological research: Decisions for scale development and validation. *Canadian Journal of Behavioural Science*, 49 (2), 78–88.

- Forman EM, Herbert JD, Moitra E, Yeomans PD, & Geller PA (2007). A randomized controlled effectiveness trial of acceptance and commitment therapy and cognitive therapy for anxiety and depression. *Behavior Modification*, 31(6), 772–799. [PubMed: 17932235]
- French K, Golijani-Moghaddam N, & Schröder T (2017). What is the evidence for the efficacy of self-help acceptance and commitment therapy? A systematic review and meta-analysis. *Journal of Contextual Behavioral Science*, 6(4), 360–374.
- Fried EI, van Borkulo CD, Epskamp S, Schoevers RA, Tuerlinckx F, & Borsboom D (2016). Measuring depression over time ... Or not? Lack of unidimensionality and longitudinal measurement invariance in four common rating scales of depression. *Psychological Assessment*, 28 (11), 1354–1367. [PubMed: 26821198]
- Friese M, & Frankenbach J (2020). P-hacking and publication bias interact to distort meta-analytic effect size estimates. *Psychological Methods*, 25(4), 456–471. [PubMed: 31789538]
- Fritz MS, & MacKinnon DP (2007). Required sample size to detect the mediated effect. *Psychological Science*, 18 (3), 233–239. 10.1111/j.1467-9280.2007.01882.x. [PubMed: 17444920]
- Gómez W, Chmielewski M, Kotov R, Ruggero C, & Watson D (2011). Development of a measure of experiential avoidance: The multidimensional experiential avoidance questionnaire. *Psychological Assessment*, 23(3), 692–713. [PubMed: 21534697]
- Gaudiano BA (2011). A review of acceptance and commitment therapy (ACT) and recommendations for continued scientific advancement. *Scientific Review of Mental Health Practice*, 8(2), 5–22.
- Gillanders DT, Bolderston H, Bond FW, Dempster M, Flaxman PE, Campbell L, Kerr S, Tansey L, Noel P, & Ferenbach C (2014). The development and initial validation of the Cognitive Fusion Questionnaire. *Behavior Therapy*, 45(1), 83–101. [PubMed: 24411117]
- Gloster AT, Block VJ, Klotsche J, Villanueva J, Rinner MT, Benoy C, Walter M, Karekla M, & Bader K (2021). Psy-Flex: A contextually sensitive measure of psychological flexibility. *Journal of Contextual Behavioral Science*, 22, 13–23.
- Gloster AT, Walder N, Levin ME, Twohig MP, & 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(6), 917–932. [PubMed: 33051034]
- Goldsmith KA, MacKinnon DP, Chalder T, White PD, Sharpe M, & Pickles A (2018). Tutorial: The practical application of longitudinal structural equation mediation models in clinical trials. *Psychological Methods*, 23(2), 191–207. 10.1037/met0000154. [PubMed: 29283590]
- Götz M, O’Boyle EH, Gonzalez-Mulé E, Banks GC, & Bollmann SS (2021). The “Goldilocks zone”: (Too) many confidence intervals in tests of mediation just exclude zero. *Psychological Bulletin*, 147(1), 95–114. [PubMed: 33382321]
- Hadash Y, & Bernstein A (2019). Behavioral assessment of mindfulness: Defining features, organizing framework, and review of emerging methods. *Current Opinion in Psychology*, 28, 229–237. [PubMed: 30959378]
- Hadash Y, Plonsker R, Vago DR, & Bernstein A (2016). Experiential self-referential and selfless processing in mindfulness and mental health: Conceptual model and implicit measurement methodology. *Psychological Assessment*, 28 (7), 856–869. [PubMed: 27078181]
- Han A, Yuen HK, Lee HY, & Zhou X (2020). Effects of acceptance and commitment therapy on process measures of family caregivers: A systematic review and meta-analysis. *Journal of Contextual Behavioral Science*, 18, 201–213.
- Harris R. (2019). ACT made simple: An easy-to-read primer on acceptance and commitment therapy. New Harbinger.
- Hayes S. (2022). ACT randomized controlled trials since 1986. Retrieved April 29, 2022, from https://contextualscience.org/act_randomized_controlled_trials_since_1986.
- Hayes SC, Hofmann SG, Stanton CE, Carpenter JK, Sanford BT, Curtiss JE, & Ciarrochi J (2019). The role of the individual in the coming era of process-based therapy. *Behaviour Research and Therapy*, 117, 40–53. [PubMed: 30348451]

- Hayes SC, Levin ME, Plumb-Villardaga J, Villatte JL, & Pistorello J (2013). Acceptance and commitment therapy and contextual behavioral science: Examining the progress of a distinctive model of behavioral and cognitive therapy. *Behavior Therapy*, 44(2), 180–198. [PubMed: 23611068]
- Hayes SC, Luoma JB, Bond FW, Masuda A, & Lillis J (2006). Acceptance and commitment therapy: Model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1–25. [PubMed: 16300724]
- Hayes SC, Merwin RM, McHugh L, Sandoz EK, A-Tjak JG, Ruiz FJ, Barnes-Holmes D, Bricker JB, Ciarrochi J, & Dixon MR (2021). Report of the ACBS Task Force on the strategies and tactics of contextual behavioral science research. *Journal of Contextual Behavioral Science*, 20, 172–183.
- Hayes SC, Strosahl K, Wilson KG, Bissett RT, Pistorello J, Toarmino D, Polusny MA, Dykstra TA, Batten SV, Bergan J, Stewart SH, Zvolensky MJ, Eifert GH, Bond FW, Forsyth JP, Karekla M, & McCurry SM (2004). Measuring experiential avoidance: A preliminary test of a working model. *Psychological Record*, 54(4), 553–578 <http://psychology.kenyon.edu/welcome.htm>.
- Hayes SC, Strosahl KD, & Wilson KG (1999). *Acceptance and commitment therapy: An experiential approach to behavior change*. Guilford Press.
- Hayes SC, Strosahl KD, & Wilson KG (2012). *Acceptance and commitment therapy: The process and practice of mindful change* (2nd ed.). Guilford Press.
- Hayes SC, 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, 141–168.
- Hofmann SG, Curtiss JE, & Hayes SC (2020). Beyond linear mediation: Toward a dynamic network approach to study treatment processes. *Clinical Psychology Review*, 76, 101824. [PubMed: 32035297]
- Johannsen M, Nissen ER, Lunderhoff M, & O’Toole MS (2022). Mediators of acceptance and mindfulness-based therapies for anxiety and depression: A systematic review and meta-analysis. *Clinical Psychology Review*, 94, 102156. [PubMed: 35483275]
- John LK, Loewenstein G, & Prelec D (2012). Measuring the prevalence of questionable research practices with incentives for truth telling. *Psychological Science*, 23(5), 524–532. [PubMed: 22508865]
- Jose PE (2016). The merits of using longitudinal mediation. *Educational Psychologist*, 51(3–4), 331–341.
- Jose PE (2019). Mediation and moderation. In Hancock GR, Stapleton LM, & Mueller RO (Eds.), *The reviewer’s guide to quantitative methods in the social sciences* (pp. 248–259). Routledge. 10.4324/9781315755649.
- Judd CM, & Kenny DA (1981). Process analysis: Estimating mediation in treatment evaluations. *Evaluation Review*, 5, 602–619.
- Kashdan TB, Goodman FR, Machell KA, Kleiman EM, Monfort SS, Ciarrochi J, & Nezlek JB (2014). A contextual approach to experiential avoidance and social anxiety: Evidence from an experimental interaction and daily interactions of people with social anxiety disorder. *Emotion*, 14(4), 769–781. [PubMed: 24749634]
- Kerr NL (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2 (3), 196–217. [PubMed: 15647155]
- Krypotos A-M, Klugkist I, Mertens G, & Engelhard IM (2019). A step-by-step guide on preregistration and effective data sharing for psychopathology research. *Journal of Abnormal Psychology*, 128(6), 517–527. [PubMed: 31368731]
- Lachowicz MJ, Preacher KJ, & Kelley K (2018). A novel measure of effect size for mediation analysis. *Psychological Methods*, 23(2), 244–261. [PubMed: 29172614]
- Landis RS, James LR, Lance CE, Pierce CA, & Rogelberg SG (2014). When is nothing something? Editorial for the null results special issue of *Journal of Business and Psychology*. *Journal of Business and Psychology*, 29(2), 163–167.
- Lilienfeld SO (2017). Psychology’s replication crisis and the grant culture: Righting the ship. *Perspectives on Psychological Science*, 12(4), 660–664. [PubMed: 28727961]

- Longmore RJ, & Worrell M (2007). Do we need to challenge thoughts in cognitive behavioral therapy? *Clinical Psychology Review*, 27, 173–187. [PubMed: 17157970]
- Lundgren T, Dahl J, & Hayes SC (2008). Evaluation of mediators of change in the treatment of epilepsy with acceptance and commitment therapy. *Journal of Behavioral Medicine*, 31(3), 225–235. [PubMed: 18320301]
- Lundgren T, Luoma JB, Dahl J, Strosahl K, & Melin L (2012). The Bull’s-Eye Values Survey: A psychometric evaluation. *Cognitive and Behavioral Practice*, 19, 518–526.
- MacKinnon DP, Fairchild AJ, & Fritz MS (2007). Mediation analysis. *Annual Review of Psychology*, 58, 593–614.
- McArdle JJ, & Grimm KJ (2010). Five steps in latent curve and latent change score modeling with longitudinal data. In van Montfort K, Oud JHL, & Satorra A (Eds.), *Longitudinal research with latent variables* (pp. 245–273). Springer. 10.1007/978-3-642-11760-2_8.
- Millsap RE (2012). *Statistical approaches to measurement invariance*. Routledge.
- Mio evi M, O’Rourke HP, MacKinnon DP, & Brown HC (2018). Statistical properties of four effect-size measures for mediation models. *Behavior Research Methods*, 50(1), 285–301. [PubMed: 28342072]
- Mokkink LB, Terwee CB, Patrick DL, Alonso J, Stratford PW, Knol DL, Bouter LM, & De Vet HC (2010). The COSMIN checklist for assessing the methodological quality of studies on measurement properties of health status measurement instruments: An international Delphi study. *Quality of Life Research*, 19(4), 539–549. [PubMed: 20169472]
- Newsome D, Newsome K, Fuller TC, & Meyer S (2019). How contextual behavioral scientists measure and report about behavior: A review of JCBS. *Journal of Contextual Behavioral Science*, 12, 347–354.
- Niles AN, Burklund LJ, Arch JJ, Lieberman MD, Saxbe DE, & Craske MG (2014). Cognitive mediators of treatment for social anxiety disorder: Comparing acceptance and commitment therapy and cognitive behavioral therapy. *Behavior Therapy*, 45(5), 664–677. [PubMed: 25022777]
- O’Neill L, Latchford G, McCracken LM, & Graham CD (2019). The development of the Acceptance and Commitment Therapy Fidelity Measure (ACT-FM): A Delphi study and field test. *Journal of Contextual Behavioral Science*, 14, 111–118.
- Ong CW, Barney JL, Barrett TS, Lee EB, Levin ME, & Twohig MP (2019). The role of psychological inflexibility and self-compassion in acceptance and commitment therapy for clinical perfectionism. *Journal of Contextual Behavioral Science*, 13, 7–16.
- Ong CW, Lee EB, Levin ME, & Twohig MP (2019). A review of AAQ variants and other context-specific measures of psychological flexibility. *Journal of Contextual Behavioral Science*, 12, 329–346.
- Ong CW, Pierce BG, Petersen JM, Barney JL, Fruge JE, Levin ME, & Twohig MP (2020). A psychometric comparison of psychological inflexibility measures: Discriminant validity and item performance. *Journal of Contextual Behavioral Science*, 18, 34–47.
- Ong CW, Pierce BG, Woods DW, Twohig MP, & Levin ME (2019c). The Acceptance and Action Questionnaire–II: An item response theory analysis. *Journal of Psychopathology and Behavioral Assessment*, 41(1), 123–134.
- Open Science Collaboration (2015). Estimating the reproducibility of psychological science. *Science*, 349(6251), aac4716. [PubMed: 26315443]
- Pavlicac JM, Schulenberg SE, & Buchanan EM (2021). Experiential Avoidance and Meaning in Life as Predictors of Valued Living: A daily diary study. *Journal of Prevention and Health Promotion*, 2(1), 135–159. 10.1177/2632077021998261.
- Piccirillo ML, & Rodebaugh TL (2019). Foundations of idiographic methods in psychology and applications for psychotherapy. *Clinical Psychology Review*, 71, 90–100. [PubMed: 30665765]
- Plumb JC, & Vilardaga R (2010). Assessing treatment integrity in acceptance and commitment therapy: Strategies and suggestions. *International Journal of Behavioral Consultation and Therapy*, 6(3), 263–295.
- Putnick DL, & Bornstein MH (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. [PubMed: 27942093]

- Reeve A, Tickle A, & Moghaddam N (2018). Are acceptance and commitment therapy-based interventions effective for reducing burnout in direct-care staff? A systematic review and meta-analysis. *Mental Health Review Journal*, 23(3), 131–155. 10.1108/MHRJ-11-2017-0052.
- Reilly ED, Ritzert TR, Scoglio AA, Mote J, Fukuda SD, Ahern ME, & Kelly MM (2019). A systematic review of values measures in acceptance and commitment therapy research. *Journal of Contextual Behavioral Science*, 12, 290–304.
- Rochefort C, Baldwin AS, & Chmielewski M (2018). Experiential avoidance: An examination of the construct validity of the AAQ-II and MEAQ. *Behavior Therapy*, 49 (3), 435–449. [PubMed: 29704971]
- Rohrer JM, Hünermund P, Arslan RC, & Elson M (2021). That’s a lot to process! Pitfalls of popular path models. *PsyArXiv. Advances in Methods and Practices in Psychological Science*, 5(2). 10.1177/25152459221095827.
- Ryff CD, Love GD, Urry HL, Muller D, Rosenkranz MA, Friedman EM, Davidson RJ, & Singer B (2006). Psychological well-being and ill-being: Do they have distinct or mirrored biological correlates? *Psychotherapy and Psychosomatics*, 75, 85–95. [PubMed: 16508343]
- Schloss HM, & Haaga DA (2011). Interrelating behavioral measures of distress tolerance with self-reported experiential avoidance. *Journal of Rational-Emotive and Cognitive-Behavior Therapy*, 29(1), 53–63. [PubMed: 21448252]
- Schoemann AM, Boulton AJ, & Short SD (2017). Determining power and sample size for simple and complex mediation models. *Social Psychological and Personality Science*, 8(4), 379–386.
- Schwalbe CS, Oh HY, & Zweben A (2014). Sustaining motivational interviewing: A meta-analysis of training studies. *Addiction*, 109(8), 1287–1294. [PubMed: 24661345]
- Sellbom M, & Tellegen A (2019). Factor analysis in psychological assessment research: Common pitfalls and recommendations. *Psychological Assessment*, 31(12), 1428–1441. [PubMed: 31120298]
- Serowik K, Khan A, LoCurto J, & Orsillo S (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.
- Simonsohn U, Nelson LD, & Simmons JP (2014). P-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science*, 9(6), 666–681. [PubMed: 26186117]
- Smout M, Davies M, Burns N, & Christie A (2014). Development of the Valuing Questionnaire (VQ). *Journal of Contextual Behavioral Science*, 3(3), 164–172.
- 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 Therapy*, 47(3), 332–362.
- Tackett JL, & Miller JD (2019). Introduction to the special section on increasing replicability, transparency, and openness in clinical psychology. *Journal of Abnormal Psychology*, 128(6), 487–492. [PubMed: 31368726]
- Terwee CB, Prinsen CA, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, De Vet HC, & Mokkink LB (2018). COSMIN methodology for evaluating the content validity of patient-reported outcome measures: A Delphi study. *Quality of Life Research*, 27(5), 1159–1170. [PubMed: 29550964]
- Thompson EM, Destree L, Albertella L, & Fontenelle LF (2021). Internet-based acceptance and commitment therapy: A transdiagnostic systematic review and meta-analysis for mental health outcomes. *Behavior Therapy*, 52 (2), 492–507. [PubMed: 33622516]
- Toth AA, Banks GC, Mellor D, O’Boyle EH, Dickson A, Davis DJ, DeHaven A, Bochantin J, & Borns J (2021). Study preregistration: An evaluation of a method for transparent reporting. *Journal of Business and Psychology*, 36(4), 553–571.
- Trompeter HR, Ten Klooster PM, Schreurs KM, Fledderus M, Westerhof GJ, & Bohlmeijer ET (2013). Measuring values and committed action with the Engaged Living Scale (ELS): Psychometric evaluation in a nonclinical sample and a chronic pain sample. *Psychological Assessment*, 25(4), 1235–1246.

- Tyndall I, Waldeck D, Pancani L, Whelan R, Roche B, & Dawson DL (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.
- Usami S, Murayama K, & Hamaker EL (2019). A unified framework of longitudinal models to examine reciprocal relations. *Psychological Methods*, 24(5), 637–657. 10.1037/met0000210. [PubMed: 30998041]
- Villatte JL, Vilaradaga R, Villatte M, Vilaradaga JCP, Atkins DC, & Hayes SC (2016). Acceptance and commitment therapy modules: Differential impact on treatment processes and outcomes. *Behaviour Research and Therapy*, 77, 52–61. [PubMed: 26716932]
- Wang W-C, Chen H-F, & Jin K-Y (2015). Item response theory models for wording effects in mixed-format scales. *Educational and Psychological Measurement*, 75(1), 157–178. [PubMed: 29795817]
- Wilson KG, Sandoz EK, 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. [PubMed: 25311291]

Table 1

Brief Summary: Challenges and Recommendations

Challenge	Recommendations
1. Distinguishing ACT processes	<ul style="list-style-type: none"> • Assess ACT-specific processes and broader process variables • Make specific predictions about process–outcome relations • Model multiple process variables simultaneously
2. Developing and validating measures of ACT processes	<ul style="list-style-type: none"> • Develop process measures that are distinct from outcome measures • Test whether process measures are ACT treatment sensitive/specific • Validate measures of interest in specific populations • Measure both trait and state versions of processes • Develop and use validated behavioral measures of ACT processes
3. Using stronger measures of ACT processes/mediators	<ul style="list-style-type: none"> • Ensure that measures used meet common psychometric soundness criteria, such as those set forth by COSMIN (Mokkink et al., 2010) • Use treatment-sensitive and specific process measures that reflect the population and content of the focal intervention • Adopt new measures quickly when they are superior to old measures • Consider creating an institutional instrument library
4. Demonstrating specificity and sensitivity	<ul style="list-style-type: none"> • Use the psychometrically soundest measures available for the process and context of interest • Match assessment schedules to theorized windows of change • Use contemporary analytic methods designed to examine change in multiple variables over time • Consider treatment fidelity’s effects on change in ACT processes
5. Performing replicable science	<ul style="list-style-type: none"> • Follow open science best practices • Perform confirmatory and replication research on ACT processes • Recruit larger samples in multisite studies or aggregate data across multiple studies that use consistent and strong process measures

Note. ACT = acceptance and commitment therapy.