

Treatment sensitivity: Its importance in the measurement of psychological flexibility

Charles Benoy^{1, 2}, Barbara Knitter², Isabell Schumann¹, Klaus Bader¹, Marc Walter¹, Andrew T. Gloster²

¹Psychiatric Hospital of the University of Basel, Switzerland

²University of Basel, Department of Psychology, Division of Clinical Psychology & Intervention Science, Switzerland

To be submitted to JCBS

(Brief empirical report = 3000 words)

Corresponding Author:

Andrew T. Gloster
University of Basel
Department of Psychology
Division of Clinical Psychology & Intervention Science
Missionsstrasse 62A
CH - 4055 Basel
Switzerland
andrew.gloster@unibas.ch

Abstract

Background: Psychological flexibility (PF) is a central construct in Acceptance and Commitment Therapy (ACT). Many studies have operationalized PF using the self-report Acceptance and Actions Questionnaire (AAQ-II). Information on the treatment sensitivity of self-report assessments of PF is lacking, however. We investigated differences in the treatment sensitivity of the AAQ-II compared to other measures of PF across various samples.

Methods: Using three different clinical samples ($N=164$), we compared the pretreatment–posttreatment change scores of the AAQ-II to those of three alternative self-report questionnaires measuring PF in a within-subject design. Sensitivity to change was assessed with effect sizes and Reliable Change Index (RC).

Results: Without exception, effect sizes and rates of clinically significant change were larger in all three alternative questionnaires and across three populations compared to the standard formulation of the AAQ-II.

Conclusions: The results of the present study show greater treatment sensitivity of three alternative questionnaires measuring PF compared to the AAQ-II. The results suggest that treatment effects concerning PF may have been underestimated depending on the wording and measure used. Implications for research on PF and ACT processes and outcomes are discussed.

Keywords: treatment sensitivity; sensitivity to change; psychological flexibility; AAQ-II; Acceptance and Action Questionnaire; psychometrics

Treatment sensitivity: Its importance in the measurement of psychological flexibility

1. Introduction

Acceptance and Commitment Therapy (ACT) is a so-called third wave behavioral psychotherapy with links to Relational Frame Theory (RFT), a basic theory on human language and cognition (Fletcher & Hayes, 2005; Hayes, Luoma, Bond, Masuda, & Lillis, 2006). The central focus of ACT is improving flexible responses to suffering, away from rigid and unhealthy coping skills towards personally important activities. This is aimed by increasing psychological flexibility (PF), the ability to stay consciously in the present moment and change to or persist in value-driven behavior (Hayes et al., 2006). Psychologically flexible behavioral patterns are important for the development, maintenance, and treatment of psychopathology (Gloster, Klotsche, Chaker, Hummel, & Hoyer, 2011; Wolgast, 2014). Furthermore, they are associated with health benefits, the etiology, maintenance, and treatment of maladaptive behavior, and various measures of psychological well-being, psychopathology, and quality of life (Chawla & Ostafin, 2007; Gloster, Meyer, & Lieb, 2017; Kashdan & Rottenberg, 2010).

The most commonly cited operationalization of PF is the Acceptance and Action Questionnaire (AAQ) (Hayes et al., 2004) and its revised version (AAQ-II) (Bond et al., 2011). The AAQ-II has demonstrated good psychometric properties, proving its factorial validity with a unitary-factor model and its construct validity with correlations and relations to similar constructs (i.e., emotional distress, life functioning, levels of depression, anxiety), and showing high internal consistency and good test-retest reliability across various samples (Bond et al., 2011; Fledderus, Oude Voshaar, ten Klooster, & Bohlmeijer, 2012; Gloster et al., 2011). At the same time, the studies validating the AAQ-II have been criticized on methodological grounds (McAndrews, Richardson, & Stopa, 2018) and the construct validity of the AAQ-II is discussed as being insufficient (Gámez, Chmielewski, Kotov, Ruggero, &

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Watson, 2011; Tyndall et al., 2018; Wolgast, 2014). Even though PF and ACT are transdiagnostic approaches (Benoy, Bader, & Schumann, 2015), some authors consider the AAQ-II not being specific enough for certain clinical subgroups and diagnosis-specific versions, such as the AAQ-OC for obsessions and compulsions, have been developed (Jacoby, Abramowitz, Buchholz, Reuman, & Blakey, 2018). In addition, the growing number of alternatives measures for PF, such as the Multidimensional Experiential Avoidance Questionnaire (MEAQ) (Gámez et al., 2011), the Comprehensive Assessment of Acceptance and Commitment Therapy Processes (CompACT) (Francis, Dawson, & Golijani-Moghaddam, 2016) or the Open and Engaged State Questionnaire (OESQ) (Benoy, Knitter, et al., 2018) underlines the importance of examining how variations of operationalizations perform in different contexts.

The AAQ-II was designed for studying the theoretical model of ACT and processes that underlie therapeutic and behavioral change, and thus its items are formulated generally (Bond et al., 2011). When not adapted and specified for clinical outcome studies, however, it could distort results for example by not measuring precisely enough changes which would lead to lacks in change sensitivity and thus underestimating treatment effects.

Treatment sensitivity, or the ability of a measure to detect treatment-related change, is an important, but relatively neglected, psychometric property for any measure used in treatment outcome or process studies. Although the AAQ is frequently used in ACT studies, we do not know of explicit tests of its treatment sensitivity.

One main factor impacting the sensitivity to change is that respondents may interpret items different as intended by the questionnaire. For example, instead of reporting actual experienced emotions, insights, or behaviors, respondents may erroneously list personal beliefs and concepts *about* such emotions, insights, or behaviors (Conner & Barrett, 2012; Robinson & Clore, 2002). Given that concepts and beliefs are more stable over time than experiences, this would impact the questionnaire's ability to detect the real change. Moreover,

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

filling out questionnaires requires multitasking (e.g. Jobe & Herrmann, 1996), a process that is influenced by various internal and external context effects (e.g. Herek & Capitanio, 1999; Sudman, Bradburn, & Schwarz, 1996). To limit these effects, the context of items must be precisely clarified as possible with wording that is unambiguous, specific, and direct (Fernandez-Ballesteros, 2004; Robie, Schmit, Ryan, & Zickar, 2000; Sudman et al., 1996; Tourangeau, 2009). This will increase the probability that respondents will judge specific instances of behavior as opposed to relying on global estimates or beliefs about their own behavior (Robinson & Clore, 2002).

According to RFT, contextual cues influence how stimuli are interpreted by respondents (Fletcher & Hayes, 2005). As the broader context of stimuli determines how they are relationally framed (Blackledge, 2003; Hayes et al., 2002), the context of item wording needs to be considered to understand, predict, and influence how respondents make judgments about events and answer questionnaires. For example, a respondent may attempt to recall their own previous behavior and then equate this with a response category “a little”, “a lot”, etc. Depending on their history with such stimuli, it is feasible that the stimulus functions of these categories transfer to the recall and thus impact the response category chosen. Respondents frame items and there are many types of relations and numbers of dimensions in this interaction between respondents and their environment/questionnaire. Thus, multiple understandings and interpretations of any single item are possible, which is why contextual cues are so important. They help respondents understand the items as the researchers intended. Including contextual cues leads to more precise measurements, which in turn enables them to reflect change more accurately.

In the AAQ-II, such wording and context specificity in certain items appear insufficient. The unspecific term *feeling* is used in multiple items, and questions are formulated generally and with few contextual cues (time, situations, behavior). Such items with lower wording specificity increase the subjective interpretations required of respondents.

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

If respondents interpret items incorrectly and report for example beliefs instead of experiences, the treatment sensitivity would be affected. By maximizing the wording specificity, one would expect to minimize interpretational confusion, measure concrete experiences, and thus increase the sensitivity to treatment-related behavioral change. Psychometrically, the treatment sensitivity would be increased.

We hypothesized that questionnaires with items that have greater wording specificity would elicit different patterns of response from questionnaires than more generally formulated items, such as the standard AAQ-II. We expected that such an increase in specificity would help respondents correctly interpret the items in the context of direct experiences. In the special case of sensitivity to change, specific treatment effects would then be more accurately measured, which in turn would lead to higher psychometric treatment sensitivity. We expected thus a greater sensitivity to change in questionnaires with higher wording specificity compared to the AAQ-II.

To examine these assumptions, we used data from three ACT intervention studies with a within-subject design and compared the AAQ-II (Bond et al., 2011; Hoyer & Gloster, 2013) to that of three different operationalizations PF with greater wording specificity an expected higher Pre-to-Post treatment effects of these questionnaires compared to the AAQ-II..

2. Method

2.1. Measures

The first alternative questionnaire we used was an expanded version of the AAQ-II, which we will further designate AAQ-II-R. In order to improve wording specificity of the AAQ-II, we simply added examples to the five items that used the term *feeling* so that they additionally read "...such as anxiety, panic, depression, etc." That is, the term *feeling* was contextualized by providing examples of the targeted emotional states. For example, AAQ-II

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Item No. 3 (“I’m afraid of my feelings”) was changed into “I’m afraid of my feelings such as anxiety, panic, depression, etc.”. The items of the AAQ-II were not otherwise changed.

The second comparison was to the Open and Engagement State Questionnaire (OESQ) (Benoy, Knitter, et al., 2018). The OESQ is a brief self-report instrument for measuring PF. Its 4 items are specific in wording, especially regarding variables of time, emotion, and behavior, for example, by providing concrete behavior-related examples and specifications related to targeted emotional states and time scales. Its main attribute appears to be its greater face validity, which it owes to its provision of a large amount of concrete information to the respondents. Its validation within three different clinical samples confirmed one-factor structure and showed good internal consistency and construct validity (Benoy, Knitter, et al., 2018).

Finally, the third comparison was to the Psyflex, a six-item self-report questionnaire assessing all six processes involved in ACT in a state form with high temporal specificity (last week) (Firsching et al., 2018). It has been constructed to be contextually sensitive (e.g., “When it is important, I can ...”) and is tailored to a low reading level to promote readability among less-educated people. It has been validated in two clinical and two community samples and appears to be a reliable and valid instrument for assessing PF. In addition to the confirmation of the one-factor-structure, it showed good discriminant and convergent validity (e.g. to other measures of PF, well-being, and psychopathology), as well as high internal consistency (Firsching et al., 2018).

2.2. Samples

Three different samples were used to analyze the treatment sensitivity of the AAQ-II. All data were obtained from clinical trials with pretreatment–posttreatment within-subject designs. Sample characteristics are described in Table 1.

The first sample was from a randomized controlled trial that tested the efficacy of ACT for treatment-resistant outpatients with primary panic disorder and/or agoraphobia (

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Gloster et al., 2015). A subset of participants was given the original version of the AAQ-II and our revised AAQ-II-R with higher wording specificity.

The second sample was a subset of participants from a randomized controlled trial that tested an ACT-based self-help online intervention for stress and burnout. In this bibliotherapeutic study, respondents had no contact with a therapist. The sample consisted of one immediate intervention group and two waiting-list control groups. All participants filled out the AAQ-II and the OESQ before and after treatment (Hofer et al., 2018). One outlier who showed strongly lower PF after treatment in the OESQ (against the hypothesis) in the first waiting-list group must be mentioned. However, to remain conservative and to protect against biasing the results in favor of our hypotheses, we decided not to exclude this outlier. When it was excluded, the differences in sensitivity to change were even slightly larger. The third sample was from an observational study about inpatient ACT-therapy for treatment-resistant mental disorders. In this sample, PF was assessed with the AAQ-II and the Psyflex and both questionnaires were filled out during the first and the last week of entering the clinic (Benoy, Meyer, et al., 2018).

2.3. Statistical procedure

Treatment sensitivity was analyzed by comparing change observed during treatment across questionnaires. Changes in the AAQ-II were compared to the respective changes in the other three questionnaires.

The significance of the difference between pre- and posttreatment was assessed for each questionnaire using paired *t*-tests. Furthermore, we used effect size measures and reliable change index (RC) to estimate sensitivity to change. This allows assessing changes both on a global/sample level (effect sizes) and on an individual level (RC). Using the RC index, the proportion of clinically significant improved subjects can be identified (Jacobson & Truax, 1991). The RC is calculated as follows: $RC = (x_{post} - x_{pre}) / SE_{rm}$, where $SE_{rm} = \sqrt{2(S_E)^2}$ and $S_E = S_1 \sqrt{1 - r_{xx}}$ (S_1 : Standard deviation of pretreatment; r_{xx} : test-retest correlation). Note that

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

in literature, the S_{ERM} (standard error of repeated measurement) is also referred to as S_{diff} (Standard Error of the differences). As test-retest correlations were not available for all measures, we used Cronbach's α as the reliability measure for all three questionnaires. We used internal consistency estimations from literature when available ($\alpha_{AAQ}=.84$, Hoyer & Gloster, 2013; $\alpha_{ESQ}=.86$, Benoy, Knitter, et al., 2018), and estimated Cronbach's alpha with pretreatment data from our sample when no prior data was available (in our sample, estimation from pretreatment data: $\alpha_{Psyflex}=.92$). When RC is greater than 1.96, the magnitude of change for a given individual is considered being statistically reliable (Jacobson & Truax, 1991). Furthermore, effect sizes were calculated as follows: $(M_{post}-M_{pre})/S_{ERM}$.

3. Results

Across all three data sets, the statistical analyses revealed a substantial difference, with all three alternative PF measurements showing higher effect sizes and higher respond rates according to RC index, This indicates better treatment sensitivity compared to the AAQ-II (Figure 1 provides a good overview). All statistical results of the sensitivity to change analyses are reported in Table 2.

Sample 1: Treatment-resistant panic disorder/agoraphobia, outpatient setting

As expected, differences in the sum scores of the AAQ-II and the AAQ-II-R all differed significantly between pre- and posttreatment. Although effect sizes on both questionnaires indicate improvements, the effect size of the AAQ-II-R is twice as high as for the AAQ-II. The difference between both questionnaires in the proportion of individuals showing significant improvement is even higher (see Table 2).

Sample 2: Burnout, online self-help setting

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Looking at all pre- to posttreatment data, changes on the AAQ-II and the OESQ were both statistically significant. Similar to results for Sample 1, the results of the sensitivity-to-change analysis showed considerably higher effect sizes and reliable change rates for the OESQ than for the AAQ-II (see Table 2).

Sample 3: Mixed diagnoses, inpatient setting

Likewise, as for sample 1 and 2, pre- to posttreatment changes were significant for both questionnaires. Once again, effect sizes and reliable change rates were about twice as high for the alternative measurement of PF (Psyflex) compared to the AAQ-II (see Table 2).

4. Discussion

We examined the treatment sensitivity of different measures of PF. Consistent with our hypotheses and without exception, effect sizes of treatment change and RC were higher in alternative measures of psychological flexibility compared to the AAQ-II across all three samples and settings, suggesting that all three alternative measures are more sensitive to treatment change than the AAQ-II. This finding is practically and clinically meaningful in that a difference of this magnitude could result in an additive effect (e.g., from a mild to moderate, moderate to large, etc.) observed in treatment trials.

In our study, by simply adding the term *feeling* in the AAQ-II (increase of wording specificity) resulted in considerable growth of the measured treatment effects. Furthermore, our results clearly show larger improvements in PF with two other validated questionnaires compared to the AAQ-II. The results are consistent across treatment settings and populations, indicating that some limitations of the AAQ-II, a widely-used PF assessment tool in ACT research, should be acknowledged. Nevertheless, and despite mentioned lacks in construct validity of the AAQ-II (Tyndall et al., 2018; Wolgast, 2014), no measure for PF can be recommended as having superior psychometric properties so far (McAndrews et al., 2018)

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

and the AAQ-II still plays a major role in research and development of ACT. It is by design broad and widely applicable, which unified and stimulated necessary research in ACT.

Although, our results indicate that changes in PF after treatment, when measured with the AAQ-II, may have been underestimated so far. Thus, mediational role of PF in psychopathology, for instance, may be even stronger than previously recognized.

Nevertheless, our study has several limitations. The main limitation is the post hoc nature of our analysis. Even if our data suggested clear results, the analyses were unplanned. Thus, while inflated type II error cannot be ignored, this concern is somewhat mitigated by the results being replicated across three samples and three questionnaires. Finally, these data do not include a test-retest period independent of the treatment (i.e., baseline assessments prior to treatment), thus it cannot be excluded that the change observed was not directly related to the treatment and that a similar amount of change would be observed without treatment. Once again, these concerns are somewhat mitigated by the fact that for two of these studies, treatment changes were significantly greater than in the control groups (Gloster et al., 2015; Hofer et al., 2018). Another limitation could be that alternative factors are responsible for the higher treatment changes. In the first sample, by mentioning specific symptoms instead of “feelings” could result in social desirability to report greater well-being, as the subject might feel pressured to indicate an improvement when specific problems are addressed. This could lead to an artificial inflation of the treatment effects. Again, this concern is also somewhat mitigated by the fact that comparable effects have been found in sample 2 and 3 with other questionnaires not mentioning specific symptoms. Further, the responsiveness to change literature suggests a different level of responsiveness depending on the length of scales. As short scales were used in this examination, we suggest further studies also need to replicate the findings with longer scales. Irrespective of the specific reason for the increased treatment sensitivity, its pattern was consistent across the three samples used in this study.

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Despite these limitations, we consider our results being relevant and meaningful. The present study reported an increase consistent with a mild to a nearly moderate effect size of pre-to-post changes in a within-subject-design when specifying items in questionnaires (by simply providing examples) or when using alternative measurements (suggesting measuring the same construct). Besides its importance for contextual behavioral sciences, these findings seem to be relevant for all psychotherapy-research-community. We furthermore consider that the present study underlines the importance of PF and that our results may suggest that the effects of PF may even be stronger than assumed so far. Nevertheless, our results should be interpreted with appropriate caution, and in addition to limitations in content validity of all measures for PF, further revisions regarding sensitivity to change needs to be addressed. Finally, interpretation of AAQ-II change scores and/or planning and working out of further studies on ACT and PF such as process measurement studies should consider the present results and add questionnaires with high wording specificity in their study-designs in order not to underestimate changes in PF.

References

- Benoy, C., Bader, K., & Schumann, I. (2015). Akzeptanz- und Commitment-Therapie : ein transdiagnostischer Ansatz. *PSYCH Up2date*, 9, 237–255.
- Benoy, C., Knitter, B., Knellwolf, L., Doering, S., Klotsche, J., & Gloster, A. T. (2018). Assessing psychological flexibility: Validation of the Open and Engaged State Questionnaire. *Journal of Contextual Behavioral Science*. <https://doi.org/10.1016/j.jcbs.2018.08.005>
- Benoy, C., Meyer, A., Knitter, B., Pinhard, K., Walter, M., Bader, K., & Gloster, A. T. (2018). Inpatient Acceptance and Commitment-Therapy for treatment-resistant disorders: an observational study. *Manuscript Submitted for Publication*.
- Blackledge, J. T. (2003). An introduction to relational frame theory: Basics and applications. *The Behavior Analyst Today*, 3(4), 421–433. <https://doi.org/10.1037/h0099997>
- Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., ... Zettle, R. D. (2011). Preliminary Psychometric Properties of the Acceptance and Action Questionnaire-II: A Revised Measure of Psychological Inflexibility and Experiential Avoidance. *Behavior Therapy*, 42, 676–688. <https://doi.org/10.1016/j.beth.2011.03.007>
- Chawla, N., & Ostafin, B. (2007). Experiential Avoidance as a Functional Dimensional Approach to Psychopathology: An Empirical Review. *Journal of Clinical Psychology*, 63(9), 871–890. <https://doi.org/10.1002/jclp>
- Conner, T. S., & Barrett, L. F. (2012). Trends in ambulatory self-report: The role of momentary experience in psychosomatic medicine. *Psychosomatic Medicine*, 74, 327–337. <https://doi.org/10.1097/PSY.0b013e3182546f18>
- Fernandez-Ballesteros, R. (2004). Self-report Questionnaires. In S. N. Haynes, E. M. Heiby, & M. Hersen (Eds.), *Comprehensive Handbook of Psychological Assessment: Volume 3: Behavioral Assessment* (pp. 194–221). Hoboken, New Jersey: John Wilnes & Sons, Inc.
- Fledderus, M., Oude Voshaar, M. A. H., ten Klooster, P. M., & Bohlmeijer, E. T. (2012). Further evaluation of the psychometric properties of the acceptance and action questionnaire-II. *Psychological Assessment*, 24(4), 925–936. <https://doi.org/10.1037/a0028200>
- Fletcher, L., & Hayes, S. C. (2005). Relational frame theory, acceptance and commitment therapy, and a functional analytic definition of mindfulness. *Journal of Rational - Emotive and Cognitive - Behavior Therapy*, 23(4), 315–336. <https://doi.org/10.1007/s10942-005-0017-7>
- Francis, A. W., Dawson, D. L., & Golijani-Moghaddam, N. (2016). The development and validation of the Comprehensive assessment of Acceptance and Commitment Therapy processes (CompACT). *Journal of Contextual Behavioral Science*, 5, 134–145. <https://doi.org/10.1016/j.jcbs.2016.05.003>
- Gámez, 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. <https://doi.org/10.1037/a0023242>
- Gloster, A. T., Klotsche, J., Chaker, S., Hummel, K. V., & Hoyer, J. (2011). Assessing Psychological Flexibility: What Does It Add Above and Beyond Existing Constructs? *Psychological Assessment*, 23(4), 970–982. <https://doi.org/10.1037/a0024135>
- Gloster, A. T., Meyer, A. H., & Lieb, R. (2017). Psychological flexibility as a malleable public health target: Evidence from a representative sample. *Journal of Contextual Behavioral Science*, 6(2), 166–171. <https://doi.org/10.1016/j.jcbs.2017.02.003>
- Gloster, A. T., Sonntag, R., Hoyer, J., Meyer, A. H., Heinze, S., Ströhle, A., ... Wittchen, H. U. (2015). Treating treatment-resistant patients with panic disorder and agoraphobia using psychotherapy: A randomized controlled switching trial. *Psychotherapy and Psychosomatics*, 84(2), 100–109. <https://doi.org/10.1159/000370162>
- Hayes, S. C., Fox, E., Gifford, E. V., Wilson, K. G., Barnes-Holmes, D., & Healy, O. (2002). Derived Relational Responding as Learned Behavior. In S. C. Hayes, D. Barnes-Holmes, & B. Roche (Eds.), *Relational Frame Theory: A post-Skinnerian account of language and cognition* (pp. 21–49). Boston, MA: Springer.
- Hayes, S. C., Luoma, J. B., Bond, F. W., Masuda, A., & Lillis, J. (2006). Acceptance and commitment therapy: model, processes and outcomes. *Behaviour Research and Therapy*, 44(1), 1–25. <https://doi.org/10.1016/j.brat.2005.06.006>
- Hayes, S. C., Strosahl, K., Wilson, K. G., Bissett, R. T., Pistorello, J., Toarmino, D., ... M, M. S. (2004). Measuring experiential avoidance: a preliminary test of a working model steven c. hayes. *The Psychological Record*, (54), 553–578.
- Herek, G. M., & Capitanio, J. P. (1999). Sex differences in how heterosexuals think about lesbians and gay men: Evidence from survey context effects. *Journal of Sex Research*, 36(4), 348–360. <https://doi.org/10.1080/00224499909552007>
- Hofer, P. D., Waadt, M., Aschwanden, R., Milidou, M., Acker, J., Meyer, A. H., ... Gloster, A. T. (2018). Self-help for stress and burnout without therapist contact: An online randomised controlled trial. *Work and*

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

- Stress*, 32(2), 189–208. <https://doi.org/10.1080/02678373.2017.1402389>
- Hoyer, J., & Gloster, A. T. (2013). Psychologische Flexibilität messen: Der Fragebogen zu Akzeptanz und Handeln II. *Verhaltenstherapie*, 23, 42–44. <https://doi.org/10.1159/000347040>
- Jacobson, N. S., & Truax, P. (1991). Clinical significance: A statistical approach to defining meaningful change in psychotherapy research. *Clinical Psychology*, 59(1), 12–19.
- Jacoby, R. J., Abramowitz, J. S., Buchholz, J., Reuman, L., & Blakey, S. M. (2018). Experiential avoidance in the context of obsessions: Development and validation of the Acceptance and Action Questionnaire for Obsessions and Compulsions. *Journal of Obsessive-Compulsive and Related Disorders*, 19, 34–43. <https://doi.org/10.1016/j.jocrd.2018.07.003>
- Jobe, J., & Herrmann, D. (1996). Implications of models of survey cognition for memory theory. In *Basic and applied memory research on practical aspects of memory* (pp. 193-205.). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, 30, 865–878. <https://doi.org/10.1016/j.cpr.2010.03.001>
- McAndrews, Z., Richardson, J., & Stopa, L. (2018). Psychometric properties of acceptance measures: A systematic review. *Journal of Contextual Behavioral Science*. <https://doi.org/10.1016/j.jcbs.2018.08.006>
- Robie, C., Schmit, M. J., Ryan, A. M., & Zickar, M. J. (2000). Effects of Item Context Specificity on the Measurement Equivalence of a Personality Inventory. *Organizational Research Methods*, 3(4), 348–365. <https://doi.org/10.1177/109442810034003>
- Robinson, M. D., & Clore, G. L. (2002). Belief and feeling: Evidence for an accessibility model of emotional self-report. *Psychological Bulletin*, 128(6), 934–960. <https://doi.org/10.1037/0033-2909.128.6.934>
- Sudman, S., Bradburn, N. M., & Schwarz, N. (1996). *Thinking About Answers: The Application of Cognitive Processes to Survey Methodology*. Jossey-Bass.
- Tourangeau, R. (2009). Remembering what happened: Memory errors and survey reports. In A. A. Stone, C. A. Bachrach, J. B. Jobe, H. S. Kurtzman, & V. S. Cain (Eds.), *The Science of Self-report: Implications for Research and Practice* (pp. 29–49). Mahwah: Lawrence Erlbaum Associates.
- Tyndall, I., Waldeck, D., Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2018). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science*, (January), 0–1. <https://doi.org/10.1016/j.jcbs.2018.09.005>
- Wolgast, M. (2014). What does the acceptance and action questionnaire (AAQ-II) really measure? *Behavior Therapy*, 45(6), 831–839. <https://doi.org/10.1016/j.beth.2014.07.002>

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Table 1

Baseline characteristics of all three samples

	Sample 1	Sample 2	Sample 3
<i>N</i> _{total}	26	95	43
Age in years, mean (SD)	36.5 ± 9.1	43.8 ± 10.1	41.2 ± 13.9
Gender, n (%)			
Female	18 (30.8)	70 (73.7)	23 (53.5)
Male	8 (30.8)	25 (26.3)	20 (46.5)
Years of Education, n (%)			
≤ 10	23 (88.5)	63 (66.3)	25 (58.1)
> 10	3 (11.5)	30 (31.6)	18 (41.9)
No formal degree	0 (0.0)	2 (2.1)	0 (0.0)
Employment, n (%)			
University Student	1 (3.8)	6 (6.3)	0 (0.0)
Job Training	1 (3.8)	0 (0.0)	0 (0.0)
Employed	16 (61.5)	84 (88.4)	17 (39.5)
Unemployed	5 (19.2)	3 (3.2)	15 (34.9)
Other	3 (11.5)	2 (2.1)	11 (25.6)
Living arrangement, n (%)			
With parents	5 (19.2)	N/A	7 (16.3)
Alone	17(65.4)	N/A	21 (48.8)
With partner	4 (15.4)	N/A	14 (32.6)
Other / Unknown	0 (0.0)	97 (100)	1 (2.3)
Marital status, n (%)			
Married	4 (15.4)	62 (65.3)	12 (27.9)
Divorced/widowed/separated	4 (15.4)	10 (10.5)	9 (20.9)
Never been married	15 (57.7)	23 (24.2)	22 (51.2)
Unknown	3 (11.5)	0 (0.0)	0 (0.0)

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Table 2

Sensitivity to Change of the AAQ-II, AAQ-II-R, OESQ, and Psyflex Among Different Settings and Samples

Table 1

Sensitivity to change of AAQ-II, AAQ-II-R, OESQ and Psyflex among different settings and samples

Sample	Setting	Questionnaire	N	Pre		Post		change score		Sig. of difference	Sensitivity of change	
				M	SD	M	SD	M _{diff}	SD _{diff}	P-value*	effect size ^a	RC ^b
PD/AG	Outpatient	AAQ-II	26	24.31	10.67	20.65	8.75	3.65	8.82	.045	-0.60	11.5%
PD/AG	Outpatient	AAQ-II-R	26	32.46	8.38	26.58	8.59	5.88	7.07	.000	-1.24	30.8%
Burnout	Online self-help	AAQ-II	95	27.22	7.92	23.12	8.01	4.11	7.11	.000	-0.92	24.2%
Burnout	Online self-help	OESQ	95	24.79	7.50	17.21	8.81	7.58	10.04	.000	-1.91	47.4%
Mixed	Inpatient	AAQ-II	43	32.28	8.88	26.60	7.89	5.67	8.39	.000	-1.13	32.6%
Mixed	Inpatient	Psyflex	43	17.84	3.75	14.63	4.13	3.21	4.75	.000	-2.14	60.5%

Note. PD/AG: Panic Disorder/Agoraphobia; *significance of difference between pre and post with paired *t*-tests; ^aeffect size: $(M_{\text{post}} - M_{\text{pre}}) / \text{SE}_{\text{rm}}$ (SE_{rm}: standard error of repeated measurement) ^bReliable Change Index: % of clients showing significant improvement

TREATMENT SENSITIVITY: ITS IMPORTANCE IN THE MEASUREMENT OF PF

Figure 1

Effect sizes of the Acceptance and Action Questionnaire II (AAQ-II) in comparison to the three alternatives in three different samples—the revised version of the AAQ-II in sample 1 (AAQ-II-R), the Open and Engaged State Questionnaire in sample 2 (OESQ), and the Psyflex questionnaire in sample 3.

