Exploring the relationship between psychological flexibility and self-report and task-based measures of cognitive flexibility

Grant, A and Cassidy, SF

http://dx.doi.org/10.1016/j.jcbs.2021.12.006

Exploring the relationship between psychological flexibility and self-report and task-based measures of cognitive flexibility

Abstract

Acceptance and Commitment Therapy (ACT) interventions aim to increase psychological flexibility (PF) leading to positive treatment outcomes. It has been suggested that successful development of PF may depend on cognitive flexibility (CF), however there is a lack of clear empirical evidence accounting for the relationship between the two constructs. The current study investigated the relationship between PF and CF in a large nonclinical sample of younger adults (*N*=246). Participants completed self-report measures of PF, and both self-report and task-based measures of CF. Results indicated that self-report, but not task-based, CF correlated with and predicted PF (r= -.49, 95% BCa CI [-.58, -.38], R^2 = .236). The results are discussed with reference to rule-governed behaviour, approaches to measurement and ACT, with suggestions for future research.

Keywords: Psychological flexibility, Cognitive flexibility, Rule-governed behaviour, Acceptance and commitment therapy

1. Introduction

Acceptance and Commitment Therapy (ACT; Hayes et al., 1999) is a contextual behavioural therapy which aims to increase psychological flexibility (PF) as the mechanism of change mediating treatment outcomes, distinguishing it from therapies which focus on symptom reduction (Hayes et al., 2006, 2011). Psychological flexibility (PF) is defined as non-judgemental acceptance of all psychological events, even if uncomfortable, with contact to the present moment and continued engagement in meaningful behaviours (Hayes et al., 2006).

The ACT model is underpinned by Relational Frame Theory (RFT). RFT provides a behaviour analytic account of language, complex cognition, and human psychological suffering based on human verbal abilities to relate stimuli and events through arbitrary relational cues not dependent on physical or formal properties of the stimuli, a concept and process referred to as arbitrarily applicable relational responding (AARR) (Dymond et al., 2017; Hayes et al., 2001; Hughes & Barnes-Holmes, 2016). While AARR and related rule-governed behaviour (RGB) offer the essential advantages of allowing humans to set and achieve goals, anticipate consequences and learn through the experience of others, RGB carries with it the risk of becoming trapped in rigid patterns, leading to human psychological suffering (Hayes et al., 1986, 1999; McAuliffe et al., 2014; Monestès et al., 2014). Verbal rules may override changing contingencies in the environment, leading to cognitive fusion and generalized pliance ('verbal dominance over behavioral regulation to the exclusion of other sources of stimulus control' and 'domination of rules followed to avoid social criticism or achieve social approval' respectively, (Hayes et al., 2013, pp. 183-186), reduced sensitivity to other sources of reinforcement (insensitivity to contingencies) and the potential for psychological problems (Hayes et al., 1986; Ruiz et al., 2019; Törneke et al., 2008). Interventions targeting problematic RGB, such as cognitive defusion, can promote psychological flexibility by viewing thoughts as events created by the mind rather than literal truths, featuring in ACT protocols when ineffective RGB acts as a barrier to living a meaningful life (Hayes et al., 1999). Such protocols can reduce the power of verbal rules, allowing behaviour to move increasingly under control of changing contingencies and values (Hayes et al., 1999). RFT defusion protocols train deictic and hierarchical relational frames, framing ongoing behaviour as 'only part of me', to reduce its discriminative functions and derive new appetitive augmental rules, behaving according to changing contingencies and values (Gil-Luciano et al., 2017; Luciano et al., 2012; Törneke et al., 2015).

The ACT model of human suffering is defined by psychological *in*flexibility (PI), involving rigid behavioural responses driven by private events such as thoughts, feelings and urges, rather than values (Hayes et al., 2006; Kashdan & Rottenberg, 2010; Stabbe et al., 2019). Generalised pliance has been implicated in PI, leading to experiential avoidance (Luciano et al., 2012; Ruiz et al., 2019; Törneke et al., 2015). PI has been proposed as a transdiagnostic process involved in the development and maintenance of psychopathology (Boulanger et al., 2010; Levin et al., 2014). Consequently, promoting PF, a cornerstone of health and wellbeing (Gloster et al., 2017; Kashdan & Rottenberg, 2010), is the target of ACT interventions.

Due to the centrality of PF in ACT treatment models, there is an imperative to gain complete understanding of the PF construct. The suggestion that executive function forms the basis for PF (Kashdan & Rottenberg, 2010) highlights one potential route to further investigate PF. Similar to PI, executive function impairment is prevalent across mental disorders and has been proposed as a transdiagnostic process linked to psychopathology

(Buckholtz & Meyer-Lindenberg, 2012; Goschke, 2014; Mathews & MacLeod, 2005; McTeague et al., 2017).

Executive functions are typically defined in neuropsychological terms, referring to higher order cognitive processes modulating lower-level cognitive functions that support abilities including flexible responding in novel situations (Banich, 2009; Diamond, 2013; Miyake et al., 2000). From an RFT perspective, executive function is defined and measured in terms of RGB, with tasks intended to capture abilities involving following or deriving rules, switching rules when necessary, and continuing with a rule in the face of behavioural control conflict (Hayes et al., 1996).

Cognitive flexibility (CF), a component of executive function, is defined as switching mental sets in response to environmental feedback, allowing flexibility in generating alternatives and changing behaviour (Dajani & Uddin, 2015; Rende, 2000). Task-shifting, or set-shifting, behavioural paradigms such as the Wisconsin Card Sorting Task (WCST; Heaton et al., 1993) are commonly used to capture CF. Perseverative errors on the task, representing behaviour maintained according to a pre-existing rule despite receiving negative feedback suggesting that the rule has changed, are indicative of cognitive *in*flexibility. Accordingly, it is suggested that the WCST captures rigid rule following reflecting insensitivity to contingencies, with previous research reporting a strong correlation between generalised pliance and cognitive *in*flexibility using the WCST (O'Connor et al., 2019).

Given the apparent conceptual overlap between PF and CF, including changing behaviour given changing contingencies, and the suggestion that executive function supports PF, the current study sought to investigate the relationship between the two constructs in a nonclinical sample of young adults, anticipating a moderate correlation.

2. Method

2.1 Sample

Participants were recruited from a nonclinical adult population via online posts to the University teaching platform, research platform Survey Circle and Facebook. Sample size was determined a priori using G*Power3 (Faul et al., 2007). To detect bivariate correlation effect size of .30 using a two tailed test, with alpha set to .05, a total of 84 participants were required to achieve 80% power. An effect size of .30 was projected based on a previous study investigating the relationship between PF and components of executive function following acquired brain injury (Whiting et al., 2017).

The total sample size of 250 participants achieved for the study deviated from initial power calculations due to higher than predicted recruitment rates. Participants who reported to be under the age of 18 years were excluded (n = 4), resulting in a final sample size of 246 participants. Sociodemographic characteristics of participants are displayed in Table 1.

[Insert Table 1 here]

2.2 Measures

The Acceptance and Action Questionnaire-II (AAQ-II) (Bond et al., 2011) was used to capture PI. The AAQ-II is a 7-item questionnaire intended to capture the broader construct of PF, targeting multiple core processes of the model (Bond et al., 2011; Hayes et al., 2006). Participants rated items such as 'My painful memories prevent me from having a fulfilling life' on a 7-point Likert scale, ranging from 1 'Never True' to 7 'Always True'. Item responses are summed to provide a total score with a theoretical range of 7-49, with higher scores representing greater PI. Previous factor analyses have established a unidimensional structure (Bond et al., 2011; Gloster et al., 2011). Internal consistency has been reported as adequate (α = .84), with test-retest reliability acceptable at 3 months (α = .81) and 12 months (α = .79)(Bond et al., 2011). In the current sample, internal consistency of the AAQ-II with coefficient omega (McDonald, 1999) was found to be excellent, ω = .91, BCa 95% CI [.89, .93].

The Cognitive Flexibility Scale (CFS) (Martin & Rubin, 1995) is a 12-item measure capturing three elements of self-report CF: the awareness of options to be flexible; willingness to be flexible; and self-efficacy in being flexible. Items including 'I have many possible ways of behaving in any given situation' are rated on a 6-point Likert scale from 1 'Strongly Disagree' to 6 'Strongly Agree'. The summed score of items provides a measure of cognitive flexibility with a theoretical range of 12-72 and with higher scores representing greater CF. The CFS has demonstrated good test-retest reliability (r = .83) over a two-week period (Martin & Rubin, 1995) and a reported one-factor structure (Dennis & Vander Wal, 2010). Reliability of the CFS in the current sample is reported for positively and negatively worded items separately as method bias appeared to affect dimensionality of the scale. Adequate internal consistency was demonstrated for both positively worded, ω = .83, BCa 95% CI [.78, .87], and negatively worded items, ω = .71, BCa 95% CI [.62, .78]. Reliability of the CFS total score is also adequate (ω = .81, BCa 95% CI [.76, .84]).

The Wisconsin Card Sorting Task (WCST) (Heaton et al., 1993). The WCST was modified and adapted for online completion, providing a task-based measure of CF. Participants were asked to match a target card with one of four key cards on each trial. The stimuli on each card varied in terms of shape, colour and number. A correct match was made when the target card matched a key card on the dimension according to the current sorting rule (shape, number or colour, see Fig. 1). Participants are told the three possible rules before the task begins, learning the correct sorting rule through trial-and-error feedback.

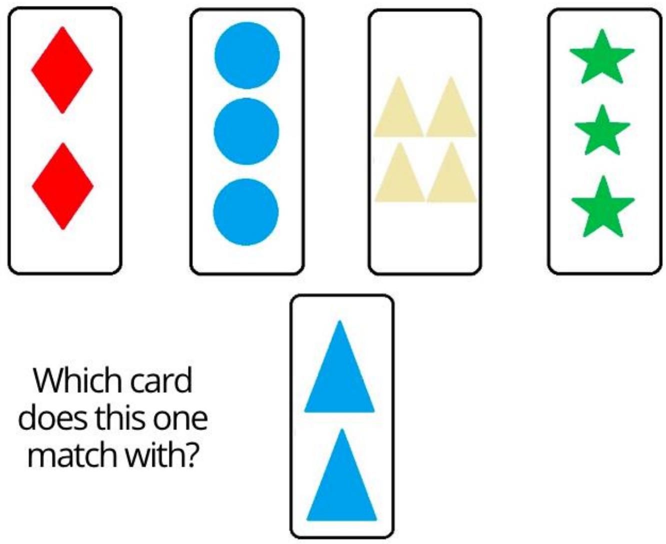

Figure 1. Card sorting task protocol

Due to inconsistencies in the literature regarding how WCST errors are defined and scored, responses on the current modified card sorting task were scored according to narrowly defined criteria described by Godinez and colleagues (2012). WCST error descriptions are displayed in Table 2.

[Insert Table 2 here]

2.3 Procedure

The study received ethics committee approval. Participants self-selected via an online recruitment post directing them to the online study hosted on Gorilla Experiment Builder. Participants were presented with participant information and consent form before progressing to demographic questions, self-report measures and WCST. The order of presentation of the self-report measures and WCST were randomised using a Latin square design. There was no opportunity to leave answers blank or to discontinue the task without withdrawing from the study, therefore there is no missing data. The study was open to participants between 28th May 2020 and 7th July 2020.

2.4 Analysis

The approach to WCST error analysis was guided by CF construct definition in the context of neurological task-based measures; *efficient errors* reflect optimal set-shifting behaviour (Barceló, 1999; Barceló & Knight, 2002), *and perseverative errors* reflect suboptimal set-shifting behaviour. Total errors were also analysed. WCST analysis focused on perseverative errors, commonly reported as the primary indicator of task-based CF (Kercood et al., 2017; Stad et al., 2019).

To mitigate minor violations of normality and outliers identified during data screening, bootstrapped confidence intervals (bias-corrected and accelerated [BCa, CI 95%,]) with 2,000 samples were computed to estimate the accuracy of correlation coefficients (Efron, 1987; Wright et al., 2011). Multiple comparisons were handled by controlling for the false discovery rate (FDR; Benjamini & Hochberg, 1995) at 5%. Age and education are known to affect performance on the WCST (Boone, 1999; Maldonado et al., 2020; Rhodes, 2004), so were controlled for in partial correlations.

2.5 Validity Checks

As the study was conducted online it is possible that some participants did not read task instructions or responded randomly when matching the target card to one of four key cards. High rates of suboptimal performance are reported in healthy research participants on common neuropsychological tests (An et al., 2012; DeRight & Jorgensen, 2015; Ross et al., 2016) indicative of low compliance and suboptimal effort that may threaten the validity of the data. To address this, separate analysis was performed on a subset of participants who achieved over 60% accuracy on the WCST ($n = 149$). The 60% accuracy threshold has been used previously to exclude participants with poor compliance on cognitive tasks (Hedge et al., 2018).

3. Results

3.1 Bivariate Correlation Analysis

Bivariate correlations with 95% bootstrapped CI (BCa) revealed a significant moderate to strong negative correlation between the AAQ-II and CFS in both the total sample ($r = -.49$, $p < .001$ [-.58, -.38]) and 60% WCST accuracy subset ($r = -.41$, $p < .001$ [-.54, -.27]). Partial correlations controlling for age and education did not alter the coefficient for the 60% WCST accuracy subset and only marginally decreased the coefficient for the total sample ($r = -.47$, $p < .001$ [-.56, -.36]).

All coefficients between the AAQ-II and WCST indices were small and non-significant ($p > 0.05$), demonstrating no meaningful relationship between AAQ-II and WSCT scores.

Small correlations were reported between the CFS and WCST errors in the total sample. Partial correlations between the CFS and total perseverative errors computed to control for age and education did not alter coefficients substantially in the total sample ($r = -.17$, $p < .01$ [-.29, -.04]) or 60% WCST accuracy subset ($r = -.20$, $p < .05$ [-.35, -.02]).

Correlations across all variables are displayed in Table 3 for the total sample, and Table 4 for the 60% WCST accuracy subset.

[Insert Table 3 here]

[Insert Table 4 here]

3.2 Regression Analysis

Linear regression was used to predict AAQ-II from CFS scores in both the total sample and 60% WCST accuracy subset. Results for the total sample indicated that the model explained 23.6% of the variance and was significant, F (1,244) = 75.17, $p < .001$, $R^2 = .236$. Self-report CF (CFS scores) predicted PI (AAQ-II scores) (unstandardised $\beta = -.572$, $p < .001$, [-.72, -.43]) (see Fig. 2).

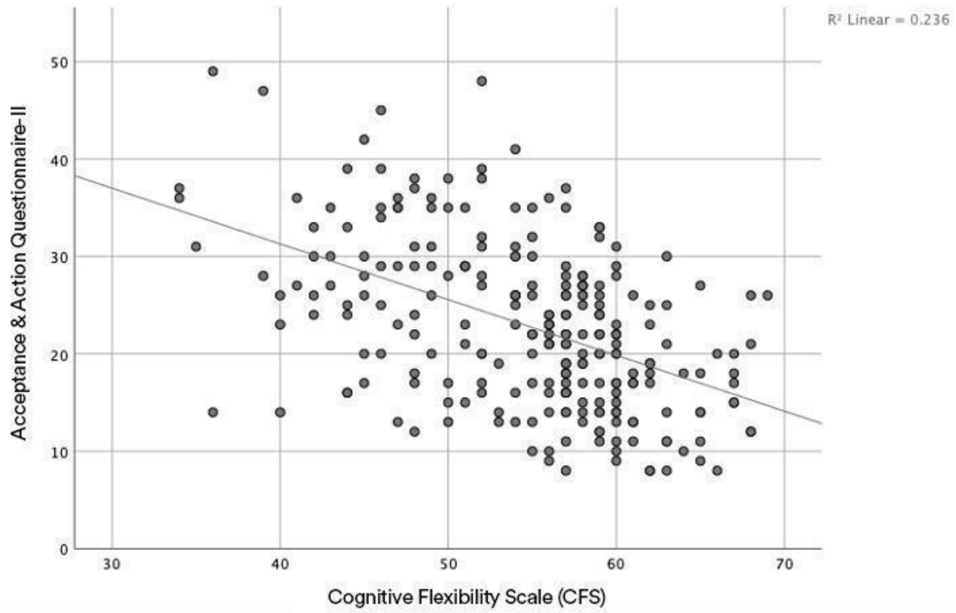

Figure 2. Scatterplot with regression line for the total sample. AAQ-II scores (intended to reflect self-report psychological inflexibility) decrease as self-report cognitive flexibility increases.

Results for the 60% WCST accuracy subset indicated that the model explained 17% of the variance and was significant, F (1,147) = 30.18, $p < .001$, $R^2 = .17$; CFS scores predicted AAQ-II scores (unstandardised $\beta = -.509$, $p < .001$, [-.71, -.32]) (see Fig. 3).

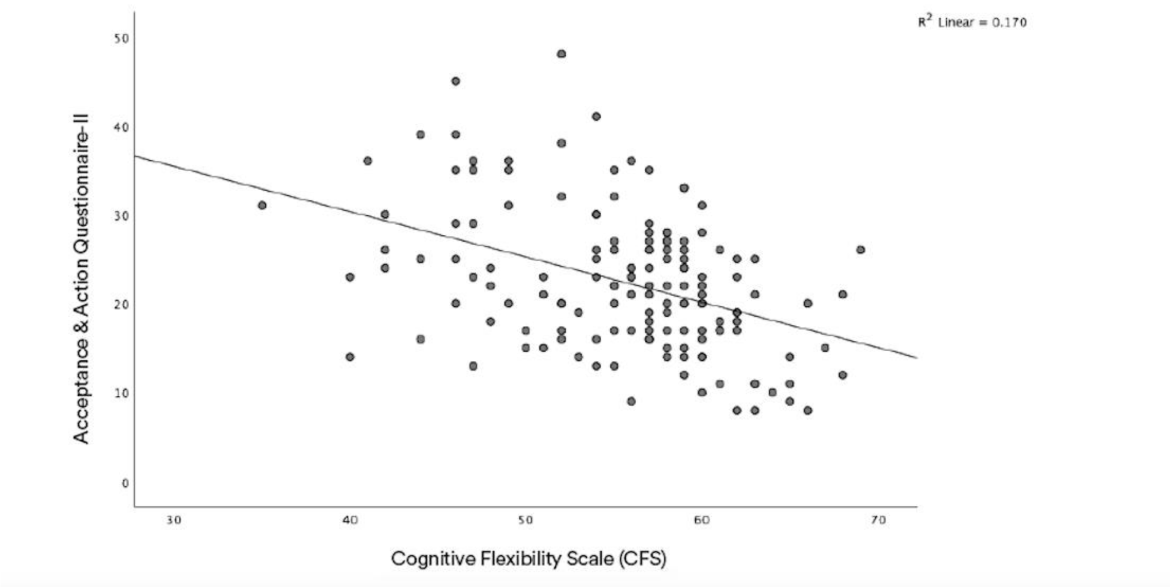

Figure 3. Scatterplot with regression line for the 60% WCST accuracy subset. AAQ-II scores (intended to capture self-report psychological inflexibility) decrease as self-report cognitive flexibility increases.

3.3 Reliabilities of WCST errors in the study sample

Reliabilities of WCST errors in the current total sample were calculated using REL_{EX}, a software tool for reliability sampling for tasks with multiple trials (Steinke & Kopp, 2020). Split-half reliability is estimated by repeatedly sampling random test splits, providing a more precise estimate than a single test split (Kopp et al., 2021; Parsons et al., 2019; Steinke & Kopp, 2020). Reliability coefficients assuming a congeneric measurement model are reported as this has the least restrictive assumptions about the data (Cho, 2016). Split-half reliability sampling using 10,000 iterations are reported in Table 3 using the Angoff-Feldt coefficient (r_{SC}) (Steinke & Kopp, 2020). The most reliable indicators of WCST performance are total errors, traditional perseverative errors and perseverative errors before a correct sort ($r_{SC} \geq .86$).

[Insert Table 5 here]

4. Discussion

To address a paucity in available research exploring the suggestion that CF is a prerequisite for the development of PF (Kashdan & Rottenberg, 2010; Whiting et al., 2017), the current study investigated the relationship between CF and PI in a nonclinical sample of younger adults using self-report and task-based measures. As anticipated, self-report CF measured using the CFS was a significant, moderate to strong, negative correlate of, and a predictor for, PI measured using the AAQ-II. However, there was no association between task-based CF, measured using the WCST, and either self-report CF or PI according to CFS and AAQ-II scores respectively.

Findings based on self-report measures are consistent with the notion of CF as a component of PF, adding credence to the suggestion that CF may be integral to successful development of PF (Whiting et al., 2017). Studies reporting an association between generalised pliance and both cognitive and psychological *in*flexibility (O'Connor et al., 2019; Ruiz et al., 2019) arguably suggest that RGB may underlie and help explain the relationship between CF and PF, based on dominance of verbal rules and insensitivity to contingencies implicated in both cognitive *in*flexibility and PI. It is worth noting that the task of exploring and explaining the relationship between CF and PI may be complicated by the fact that the processes originate in different theoretical models; the cognitivist view of CF emphasises the role of an internal mediating system, whereas PI developed from an RFT perspective emphasising the functional properties of verbal events, where such events are considered activities rather than products (Gross & Fox, 2009).

The lack of association between task-based CF, measured using the WCST, and self-report CF and PI is not unexpected. Firstly, a distinction between self-report and neuropsychological task-based measures of CF, such as the WCST, is evident in the extant literature. Different aspects of CF are captured depending on the selected approach, subjective self-report or objective task based. This distinction is supported by evidence demonstrating poor convergent validity between a range of neuropsychological measures and two separate self-report measures of CF, including the CFS (Johnco, Wuthrich & Rapee, 2014) and reports of a lack of association between CFS scores and perseverative errors on the WCST (Kercood et al., 2017). Comparable inconsistencies have also been identified in the wider executive function literature (Toplak et al., 2013). Task-based neuropsychological measures of CF have been described as 'reductionist', capturing optimal performance, whereas self-report measures capture a cumulative appraisal of typical behaviour (Wennerhold & Friese, 2020; Whiting et al., 2017). Findings from the current study indicate that the aspect of CF captured by self-report measures may be most relevant to PF measured using the AAQ-II, and therefore ACT.

4.1. Limitations and Future Research

The limitations of the study's cross sectional correlation design with regards to causal inferences, and reliance on self-report measures, should be considered when interpreting the findings. Future study designs will benefit from longitudinal aspects, with multiple data points enabling predictive conclusions, and should include behavioural measures capturing PF (cf. Gil-Luciano et al., 2017; López-López & Luciano, 2017) to provide insight into the relationship between behavioural task-based and self-report measures of PF.

Finally, despite extensive use as a measure of PF and indicator of ACT effectiveness, the construct validity of the AAQ-II has been questioned. PF is proposed as a process which promotes resilience and wellbeing (Hayes et al., 2006; Kashdan et al., 2020; Kashdan & Rottenberg, 2010). However, studies exploring its psychometric properties suggest that the

AAQ-II is capturing negative affect, psychological distress, or neuroticism, so is measuring the outcome of PI, not the process of values-consistent behaviour in the presence of uncomfortable internal experiences (Rochefort et al., 2018; Tyndall et al., 2019; Wolgast, 2014). Including additional PF measures will allow future studies to gather evidence regarding the validity of the AAQ-II as a measure of PF and help inform interpretation of existing evidence regarding the relationship between PF and CF.

5. Conclusions

The need for an increased understanding of the nature of the relationship between psychological and cognitive flexibility and potential implications for clinical practice have previously been highlighted (Whiting et al., 2017). The study provides evidence for an association between PI, measured using the AAQ-II, and self-report CF, measured using the CFS, with RGB suggested as a possible explanation forming the basis to the association. The lack of association reported between AAQ-II or CFS scores and the WCST supports the assertion that self-report and neuropsychological task-based measures may capture different aspects of CF, leading to different outcomes in studies depending on the approach to measurement. This issue also needs to be explored in relation to PF, with study designs including both self-report and behavioural measures.

From a clinical perspective, findings imply that considering impairments in self-report CF may be valuable in cases of treatment interventions such as ACT, where increased PF is a goal, warranting further investigation.

References

An, K. Y., Zakzanis, K. K., & Joordens, S. (2012). Conducting research with non-clinical healthy undergraduates: Does effort play a role in neuropsychological test performance? *Archives of Clinical Neuropsychology*, *27*(8), 849–857. https://doi.org/10.1093/arclin/acs085

Banich, M. T. (2009). Executive Function: The Search for an Integrated Account. *Current Directions in Psychological Science*, *18*(2), 89–94. https://doi.org/10.1111/j.1467-8721.2009.01615.x

Barceló, F. (1999). Electrophysiological evidence of two different types of error in the Wisconsin Card Sorting Test. *NeuroReport: For Rapid Communication of Neuroscience Research*, *10*(6), 1299–1303. https://doi.org/10.1097/00001756-199904260-00027

Barceló, F., & Knight, R. T. (2002). Both random and perseverative errors underlie WCST deficits in prefrontal patients. *Neuropsychologia*, *40*(3), 349–356. https://doi.org/10.1016/s0028-3932(01)00110-5

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289–300. JSTOR.

Bond, F. W., Hayes, S. C., Baer, R. A., Carpenter, K. M., Guenole, N., Orcutt, H. K., Waltz, T., & Zettle, R. D. (2011). Preliminary psychometric properties of the Acceptance and Action Questionnaire-II: A revised measure of psychological inflexibility and experiential avoidance. *Behavior Therapy*, *42*(4), 676–688. https://doi.org/10.1016/j.beth.2011.03.007

Boone, K. B. (1999). Neuropsychological assessment of executive functions: Impact of age, education, gender, intellectual level, and vascular status on executive test scores. In B. L. Miller & J. L. Cummings (Eds.), *The human frontal lobes: Functions and disorders* (pp. 247–260). The Guilford Press.

Boulanger, J. L., Hayes, S. C., & Pistorello, J. (2010). Experiential avoidance as a functional contextual concept. In A. M. Kring & D. M. Sloan (Eds.), *Emotion regulation and psychopathology: A transdiagnostic approach to etiology and treatment* (pp. 107–136). The Guilford Press.

Buckholtz, J. W., & Meyer-Lindenberg, A. (2012). Psychopathology and the human connectome: Toward a transdiagnostic model of risk for mental illness. *Neuron*, *74*(6), 990–1004. https://doi.org/10.1016/j.neuron.2012.06.002

Cho, E. (2016). Making Reliability Reliable: A Systematic Approach to Reliability Coefficients. *Organizational Research Methods, 19*(4), 651-682. https://doi.org/10.1177/1094428116656239

Dajani, D. R., & Uddin, L. Q. (2015). Demystifying cognitive flexibility: Implications for clinical and developmental neuroscience. *Trends in Neurosciences*, *38*(9), 571–578. https://doi.org/10.1016/j.tins.2015.07.003

Dennis, J., & Vander Wal, J. (2010). The Cognitive Flexibility Inventory: Instrument Development and Estimates of Reliability and Validity. *Cognitive Therapy and Research*, *34*(3), 241–253. https://doi.org/10.1007/s10608-009-9276-4

DeRight, J., & Jorgensen, R. S. (2015). I Just Want My Research Credit: Frequency of Suboptimal Effort in a Non-Clinical Healthy Undergraduate Sample. *The Clinical Neuropsychologist*, *29*(1), 101–117. https://doi.org/10.1080/13854046.2014.989267

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, *64*(1), 135–168. https://doi.org/10.1146/annurev-psych-113011-143750

Dymond, S., Bennett, M., Boyle, S., Roche, B., & Schlund, M. (2017). Related to Anxiety: Arbitrarily Applicable Relational Responding and Experimental Psychopathology Research on Fear and Avoidance. *Perspectives on Behavior Science*, *41*(1), 189–213. https://doi.org/10.1007/s40614-017-0133-6

Efron, B. (1987). Better Bootstrap Confidence Intervals. *Journal of the American Statistical Association*, *82*(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Gil-Luciano, B., Ruiz, F. J., Valdivia-Salas, S., & Suárez-Falcón, J. C. (2017). Promoting psychological flexibility on tolerance tasks: Framing behavior through deictic/hierarchical relations and specifying augmental functions. *The Psychological Record*, *67*(1), 1–9. https://doi.org/10.1007/s40732-016-0200-5

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

Godinez, D. A., Friedman, N. P., Rhee, S. H., Miyake, A., & Hewitt, J. K. (2012). Phenotypic and Genetic Analyses of the Wisconsin Card Sort. *Behavior Genetics*, *42*(2), 209–220. https://doi.org/10.1007/s10519-011-9502-1

Goschke, T. (2014). Dysfunctions of decision-making and cognitive control as transdiagnostic mechanisms of mental disorders: Advances, gaps, and needs in current research.

International Journal of Methods in Psychiatric Research, *23*(Suppl 1), 41–57. https://doi.org/10.1002/mpr.1410

Gross, A. C., & Fox, E. J. (2009). Relational frame theory: an overview of the controversy. The Analysis of verbal behavior, 25(1), 87–98. https://doi.org/10.1007/BF03393073

Hayes, S. C., Barnes-Holmes, D., & Roche, B. (Eds.). (2001). *Relational Frame Theory: A Post-Skinnerian Account of Human Language and Cognition*. Kluwer Academic. https://doi.org/10.1007/b108413

Hayes, S. C., Brownstein, A. J., Zettle, R. D., Rosenfarb, I., & Korn, Z. (1986). Rule-governed behavior and sensitivity to changing consequences of responding. *Journal of the Experimental Analysis of Behavior*, *45*(3), 237–256. https://doi.org/10.1901/jeab.1986.45-237

Hayes, S. C., Gifford, E. V., & Ruckstuhl Jr., L. E. (1996). Relational frame theory and executive function: A behavioral approach. In G. R. Lyon & N. A. Krasnegor (Eds.), *Attention, Memory, and Executive Function* (pp. 297–305). Paul Brookes Publishing Co.

Hayes, S. C., Levin, M. E., Plumb-Vilardaga, J., Villatte, J. L., & 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. https://doi.org/10.1016/j.beth.2009.08.002

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. (1999). *Acceptance and Commitment Therapy: An Experiential Approach to Behavior Change*. Guildford Press.

Hayes, S. C., Villatte, M., Levin, M., & Hildebrandt, M. (2011). Open, Aware, and Active: Contextual Approaches as an Emerging Trend in the Behavioral and Cognitive Therapies. *Annual Review of Clinical Psychology, 7*(1), 141-168. https://doi.org/10.1146/annurev-clinpsy-032210-104449

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtis, G. (1993). *Wisconsin card sorting test manual: Revisited and expanded*. Psychological Assessment Resources.

Hedge, C., Powell, G., & Sumner, P. (2018). The reliability paradox: Why robust cognitive tasks do not produce reliable individual differences. *Behavior Research Methods*, *50*(3), 1166–1186. https://doi.org/10.3758/s13428-017-0935-1

Hulbert-Williams, L., Hochard, K., Hulbert-Williams, N., Archer, R., Nicholls., W. & Wilson., K. (2016). Contextual behavioural coaching: An evidence-based model for supporting behaviour change. *International Coaching Psychology Review, 11*(2), 142-154.

Hughes, S., & Barnes-Holmes, D. (2016). Relational frame theory: The basic account. In *The Wiley handbook of contextual behavioral science* (pp. 129–178). Wiley Blackwell.

Johnco, C., Wuthrich, V. M., & Rapee, R. M. (2014). Reliability and validity of two self-report measures of cognitive flexibility. *Psychological Assessment*, *26*(4), 1381–1387. https://doi.org/10.1037/a0038009

Kashdan, T. B., Disabato, D. J., Goodman, F. R., Doorley, J. D., & McKnight, P. E. (2020). Understanding psychological flexibility: A multimethod exploration of pursuing valued goals despite the presence of distress. *Psychological Assessment*, *32*(9), 829–850. https://doi.org/10.1037/pas0000834

Kashdan, T. B., & Rottenberg, J. (2010). Psychological flexibility as a fundamental aspect of health. *Clinical Psychology Review*, *30*(7), 865–878. https://doi.org/10.1016/j.cpr.2010.03.001

Kercood, S., Lineweaver, T. T., Frank, C. C., & Fromm, E. D. (2017). Cognitive Flexibility and Its Relationship to Academic Achievement and Career Choice of College Students with and without Attention Deficit Hyperactivity Disorder. *Journal of Postsecondary Education and Disability*, *30*(4), 329–344.

Kopp, B., Lange, F., & Steinke, A. (2021). The Reliability of the Wisconsin Card Sorting Test in Clinical Practice. *Assessment*, *28*(1), 248-263 https://doi.org/10.1177/1073191119866257

Levin, M. E., MacLane, C., Daflos, S., Seeley, J., Hayes, S. C., Biglan, A., & Pistorello, J. (2014). Examining psychological inflexibility as a transdiagnostic process across psychological disorders. *Journal of Contextual Behavioral Science*, *3*(3), 155–163. https://doi.org/10.1016/j.jcbs.2014.06.003

López-López, J. C., & Luciano, C. (2017). An experimental analysis of defusion interactions based on deictic and hierarchical framings and their impact on cognitive performance. *The Psychological Record*, *67*(4), 485–497. https://doi.org/10.1007/s40732-017-0250-3

Luciano, C., Valdivia-Salas, S., & Ruiz, F. J. (2012). The self as the context for rule-governed behavior. In L. McHugh & I. Stewart (Eds.), *The self and perspective taking: Contributions and applications from modern behavioral science* (pp. 143–160). New Harbinger Publications.

Maldonado, T., Orr, J. M., Goen, J. R. M., & Bernard, J. A. (2020). Age Differences in the Subcomponents of Executive Functioning. *The Journals of Gerontology: Series B*, *75*(6), e31–e55. https://doi.org/10.1093/geronb/gbaa005

Martin, M. M., & Rubin, R. B. (1995). A New Measure of Cognitive Flexibility. *Psychological Reports*, *76*(2), 623–626. https://doi.org/10.2466/pr0.1995.76.2.623

Mathews, A., & MacLeod, C. (2005). Cognitive Vulnerability to Emotional Disorders. *Annual Review of Clinical Psychology*, *1*(1), 167–195. https://doi.org/10.1146/annurev.clinpsy.1.102803.143916

McAuliffe, D., Hughes, S., & Barnes-Holmes, D. (2014). The Dark-Side of Rule Governed Behavior: An Experimental Analysis of Problematic Rule-Following in an Adolescent Population With Depressive Symptomatology. *Behavior Modification*, *38*(4), 587–613. https://doi.org/10.1177/0145445514521630

McDonald, R. P. (1999). *Test theory: A unified treatment* (pp. xi, 485). Lawrence Erlbaum Associates Publishers.

McTeague, L. M., Huemer, J., Carreon, D. M., Jiang, Y., Eickhoff, S. B., & Etkin, A. (2017). Identification of Common Neural Circuit Disruptions in Cognitive Control Across Psychiatric Disorders. *American Journal of Psychiatry*, *174*(7), 676–685. https://doi.org/10.1176/appi.ajp.2017.16040400

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex “Frontal Lobe” Tasks: A Latent Variable Analysis. *Cognitive Psychology*, *41*(1), 49–100. https://doi.org/10.1006/cogp.1999.0734

Monestès, J. L., Villatte, M., Stewart, I., & Loas, G. (2014). Rule-Based Insensitivity and Delusion Maintenance in Schizophrenia. *The Psychological Record*, *64*(2), 329–338. https://doi.org/10.1007/s40732-014-0029-8

O’Connor, M., Byrne, P., Ruiz, F. J., & McHugh, L. (2019). Generalized Pliance in Relation to Contingency Insensitivity and Mindfulness. *Mindfulness*, *10*(5), 833–840. https://doi.org/10.1007/s12671-018-1046-5

Parsons, S., Kruijt, A.-W., & Fox, E. (2019). Psychological Science Needs a Standard Practice of Reporting the Reliability of Cognitive-Behavioral Measurements. *Advances in Methods and Practices in Psychological Science*, *2*(4), 378–395. https://doi.org/10.1177/2515245919879695

Rende, B. (2000). COGNITIVE FLEXIBILITY: THEORY, ASSESSMENT, AND TREATMENT. *Seminars in Speech and Language*, *21*(2), 121–153. https://doi.org/10.1055/s-2000-7560

Rhodes, M. G. (2004). Age-related differences in performance on the Wisconsin card sorting test: A meta-analytic review. *Psychology and Aging*, *19*(3), 482–494. https://doi.org/10.1037/0882-7974.19.3.482

Rochefort, C., Baldwin, A. S., & Chmielewski, M. (2018). Experiential Avoidance: An Examination of the Construct Validity of the AAQ-II and MEAQ. *Behavior Therapy*, *49*(3), 435–449. https://doi.org/10.1016/j.beth.2017.08.008

Ross, T. P., Poston, A. M., Rein, P. A., Salvatore, A. N., Wills, N. L., & York, T. M. (2016). Performance Invalidity Base Rates Among Healthy Undergraduate Research Participants. *Archives of Clinical Neuropsychology*, *31*(1), 97–104. https://doi.org/10.1093/arclin/acv062

Ruiz, F. J., Suárez-Falcón, J. C., Barbero-Rubio, A., & Flórez, C. L. (2019). Development and initial validation of the Generalized Pliance Questionnaire. *Journal of Contextual Behavioral Science*, *12*, 189–198. https://doi.org/10.1016/j.jcbs.2018.03.003

Stabbe, O. K., Rolffs, J. L., & Rogge, R. D. (2019). Flexibly and/or inflexibly embracing life: Identifying fundamental approaches to life with latent profile analyses on the dimensions of the Hexaflex model. *Journal of Contextual Behavioral Science*, *12*, 106–118. https://doi.org/10.1016/j.jcbs.2019.03.003

Stad, F. E., Wiedl, K. H., Vogelaar, B., Bakker, M., & Resing, W. C. M. (2019). The role of cognitive flexibility in young children's potential for learning under dynamic testing conditions. *European Journal of Psychology of Education*, *34*(1), 123–146. https://doi.org/10.1007/s10212-018-0379-8

Steinke, A., & Kopp, B. (2020). RELEX: An Excel-based software tool for sampling split-half reliability coefficients. *Methods in Psychology*, *2*, 100023. https://doi.org/10.1016/j.metip.2020.100023

Toplak, M. E., West, R. F., & Stanovich, K. E. (2013). Practitioner review: do performance-based measures and ratings of executive function assess the same construct?. *Journal of child psychology and psychiatry, 54(*2), 131–143. https://doi.org/10.1111/jcpp.12001

Törneke, N., Luciano, C., Barnes-Holmes, Y., & Bond, F. W. (2015). RFT for Clinical Practice: Three Core Strategies in Understanding and Treating Human Suffering. In R. D. Zettle, S. C. Hayes, D. Barnes-Holmes, & A. Biglan (Eds.), *The Wiley Handbook of Contextual Behavioral Science* (pp. 254–273). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118489857.ch12

Törneke, N., Luciano, C., & Salas, S. V. (2008). Rule-governed behavior and psychological problems. *International Journal of Psychology & Psychological Therapy*, *8*(2), 141–156.

Tyndall, I., Waldeck, D., Pancani, L., Whelan, R., Roche, B., & Dawson, D. L. (2019). The Acceptance and Action Questionnaire-II (AAQ-II) as a measure of experiential avoidance: Concerns over discriminant validity. *Journal of Contextual Behavioral Science*, *12*, 278–284. https://doi.org/10.1016/j.jcbs.2018.09.005

Wennerhold, L., & Friese, M. (2020). Why Self-Report Measures of Self-Control and

Inhibition Tasks Do Not Substantially Correlate. *Collabra: Psychology*, *6*(1), 9. https://doi.org/10.1525/collabra.276

Whiting, D. L., Deane, F. P., Simpson, G. K., Mcleod, H. J., & Ciarrochi, J. (2017). Cognitive and psychological flexibility after a traumatic brain injury and the implications for treatment in acceptance-based therapies: A conceptual review. *Neuropsychological Rehabilitation*, *27*(2), 263–299. https://doi.org/10.1080/09602011.2015.1062115

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

Wright, D. B., London, K., & Field, A. P. (2011). Using Bootstrap Estimation and the Plug-in Principle for Clinical Psychology Data. *Journal of Experimental Psychopathology*, *2*(2), 252–270. https://doi.org/10.5127/jep.013611

Table 1

Sociodemographic characteristics of participants

Sociodemographic Characteristics (*N* = 246)	n	%
Gender		
Female	167	67.9
Male	79	32.1
Ethnicity		
White	194	78.9
Asian/Asian British	31	12.6
Mixed/Multiple Ethnic Groups	11	4.5
Black/African/Caribbean/Black British	3	1.2
Arab	3	1.2
Hispanic/Latino	2	0.8
Middle Eastern	1	0.4
Indian	1	0.4
Education Level		
Degree level and above	221	89.8
2+ A-Levels or Equivalent (e.g., NVQ Level 3)	12	4.9
5+ GCSE's or Equivalent (e.g., NVQ Level 2)	6	2.4
1-4 GCSE's or Equivalent (e.g., NVQ Level 1)	3	1.2
Apprenticeship	2	0.8
No Qualifications	3	1.2

Note. *N* = 246. Participant age range was 18 to 67 years (*M* = 30.18, *SD* = 8.42).

Table 2

Description of WCST Error Scoring Indices

Error Type	Description
Total Errors	Errors made throughout all blocks were scored and summed
Total Perseverative Errors	A participant continued to sort according to the previously correct rule after receiving feedback that the rule had now changed
Perseverative Errors Before Correct Sort	A participant continues to sort according to the previously correct rule after receiving feedback that the rule had now changed, before identifying the new sorting rule
Perseverative Errors After Correct Sort	A participant reverts back to sort according to the previously correct rule after correctly switching to the new rule
Efficient Error	A participant receives negative feedback following a rule change and subsequently switches to the correct rule on the following trial, reflecting optimal shifting behaviour

Table 3

Bivariate correlation coefficients for psychological flexibility and cognitive flexibility for total sample

Variable	1	2	3	4	5	6	7	8
1. Age								
2. Education	-.10							
	[-.25, .05]							
3. AAQ-II	-.12	-.08						
	[-.24, .01]	[-.25, .09]						
4. CFS	.26**	.12	-.49**					
	[.15, .36]	[-.03, .26]	[-.58, -.38]					
5. Total	-.14	-.12	.07	-.17*				
Errors	[-.24, .04]	[-.28, .04]	[-.05, .20]	[-.30, -.05]				
6. Total PE	-.15	-.10	.11	-.22*	.86**			
	[-.25, -.04]	[-.27, .05]	[-.01, .23]	[-.33, -.09]	[.83, .89]			
7. PE BCS	-.11	.10	.10	-.18*	.72**	.91**		
	[-.21, .01]	[-.28, .06]	[-.02, .22]	[-.30, -.06]	[.66, .79]	[.86, .94]		
8. PE ACS	-.12	-.04	.04	-.13	.55**	.50**	.09	
	[-.21, -.02]	[-.18, .07]	[-.10, .21]	[-.25, .004]	[.46, .63]	[.37, .62]	[.02, .21]	
9. Efficient	.11	.07	-.06	.12	-.82**	-.81**	-.75**	-.38**
errors	[.01, .23]	[-.05, .19]	[-.18, .06]	[.002, .26]	[-.86, -.77]	[-.84, -.82]	[-.80, -.70]	[-.47, -.30]

Note. Correlation coefficients (2-tailed) are presented for the total sample (*N* = 246). 95% bootstrapped confidence intervals (BCa, 2000 samples) are presented in square brackets. PE = Perseverative error; BCS = Before correct sort; ACS = After correct sort

*Correlations significant at the .01 level, survived controlling the false discovery rate (FDR) at 5%.

**Correlations $p < .001$, survived controlling the false discovery rate (FDR) at 5%

Table 4

Bivariate correlation coefficients for psychological flexibility and cognitive flexibility for 60% WCST accuracy subset

Variable	1	2	3	4	5	6	7	8
1. Age								
2. Education	-.19							
	[-.36, .01]							
3. AAQ-II	-.09	.02						
	[-.25, .09]	[-.12, .18]						
4. CFS	.23*	.02	-.41**					
	[.08, .36]	[-.14, .14]	[-.54, -.27]					
5. Total	.15	.05	-.04	.03				
Errors	[.02, .29]	[-.04, .12]	[-.23, .15]	[-.19, .13]				
6. Total PE	.12	.04	.13	-.16	.79**			
	[-.04, .28]	[-.03, .10]	[-.08, .31]	[-.32, .01]	[.73, .85]			
7. PE BCS	.12	.03	.06	-.07	.72**	.90**		
	[-.05, .31]	[-.05, .10]	[-.12, .24]	[-.23, .09]	[.63, .79]	[.84, .93]		
8. PE ACS	.03	.05	.16	-.21*	.40**	.53**	.10	
	[-.11, .18]	[-.06, .14]	[-.02, .33]	[-.36, .05]	[.27, .52]	[.38, .66]	[-.10, 30]	
9. Efficient	-.10	-.04	.03	.02	-.67**	-.74**	-.81**	-.20
errors	[-.24, .05]	[-.12, .05]	[-.12, .18]	[-.14, .18]	[-.76, -.58]	[-.81, -.66]	[-.87, -.74]	[-.29, .05]

Note. Correlation coefficients (2-tailed) are presented for participants who achieved at least 60% accuracy on the WCST (n = 149). 95% bootstrapped confidence intervals (BCa, 2000 samples) are presented in square brackets. PE = Perseverative error; BCS = Before correct sort; ACS = After correct sort

*Correlations significant at the .01 level, survived controlling the false discovery rate (FDR) at 5%.

**Correlations p < .001, survived controlling the false discovery rate (FDR) at 5%

Table 5

Reliability of WCST errors types reported for the total study sample

		95% HDI	
	Median	Low	High
Total Errors	.92	.87	.95
Traditional PE	.86	.80	.90
PE (Before Correct Sort)	.87	.80	.91
PE (After Correct Sort)	.73	.63	.80
Efficient Errors	.47	.20	.70

Note. WCST = Wisconsin Card Sorting Task; PE = Perseverative Error; HDI = Highest Density Interval (low = lower limit at 95% credibility; high = higher limit at 95% credibility). Median and HDI of Angoff-Feldt coefficients from 10,000 iterations are reported. Scores from the total sample are reported (*N* = 246)