



Published in final edited form as:

Curr Dir Psychol Sci. 2022 June ; 31(3): 262–271. doi:10.1177/09637214221091824.

Do Rating and Task Measures of Control Abilities Assess the Same Thing?

Naomi P. Friedman¹,

Institute for Behavioral Genetics and Department of Psychology and Neuroscience, University of Colorado Boulder

Daniel E. Gustavson

Department of Medicine and Vanderbilt Genetics Institute, Vanderbilt University Medical Center

Abstract

The ability to control our thoughts and actions is broadly associated with health and success, so it is unsurprising that measuring self-control abilities is a common goal across many areas of psychology. Puzzlingly, however, different measures of control — questionnaire ratings and computerized cognitive tasks — show only weak relationships to each other. We review evidence that this discrepancy is not just a result of poor reliability or validity of ratings or tasks. Rather, ratings and tasks seem to assess different aspects of control, distinguishable along six main dimensions. To improve the psychological science surrounding self-control, it will be important for future work to investigate the relative importance of these dimensions to the dissociations between self-control measures, and for researchers to motivate and explain which aspects of control they are studying when one or both types of measures are deployed. Keywords: cognitive control, executive function, self-regulation, self-control, impulsivity

A typical day includes numerous obstacles to achieving our goals. We might have to make ourselves start a difficult project, juggle multiple tasks, ignore distractions, resist temptations, and/or tamp down strong emotions. Overcoming these challenges requires controlling our thoughts and actions. Individual differences in such control abilities are associated with numerous health and success outcomes (Diamond, 2013; Sharma et al., 2014), so it is unsurprising that they have been extensively studied. Within the psychological literature, there are two major approaches to measuring control abilities. The first involves “self-control” or “self-regulation” questionnaires that ask participants (or their families/friends) to rate their typical behavior when confronted with certain challenges (henceforth referred to as “ratings”; Table 1). The second involves laboratory “cognitive control” or “executive function” tasks (henceforth referred to as “tasks”; Fig. 1).

Ratings and tasks are often discussed as alternative measures of conceptually similar control processes, such as stopping inappropriate responses. Rating scales get at individual differences in these processes through subjective reports of typical behavior across multiple contexts, whereas tasks provide a more objective window into the cognitive mechanisms

¹Address correspondence to Naomi Friedman, Institute for Behavioral Genetics, 447 UCB, University of Colorado Boulder, Boulder, CO 80309. naomi.friedman@colorado.edu, Phone: 303-735-4457, Fax: 303-492-8063.

that may underlie these behaviors. Surprisingly, however, ample evidence now suggests that tasks and ratings do not correlate well with each other (Cyders & Coskunpinar, 2011; Sharma et al., 2014). For example, a meta-analysis (Duckworth & Kern, 2011) found that the average correlation between executive function tasks and self-report ratings was only $r=.10$. Although significantly larger than zero, this correlation indicates only a small overlap. In contrast, tasks or ratings typically correlate well ($r=.3-.5$) within domain (e.g., between working memory tasks or between lack of premeditation scales), though there are exceptions (inhibition tasks typically poorly correlate with each other; Friedman & Miyake, 2004).

In this article, we consider two primary explanations for this surprisingly low overlap of tasks and ratings and discuss their implications for the psychological science of self-control: 1) tasks and/or ratings do not reliably measure control, or 2) tasks and ratings measure different aspects of control. We argue that both ratings and tasks can be valid and reliable measures of control, but that they assess different — both meaningful — aspects of control. Clarifying these differences will be important for the design and interpretation of studies of control.

Can poor measurement explain the low correlation between ratings and tasks?

Reliability is the extent to which a measure correlates with itself (e.g., when repeatedly administered). When interpreting correlations, knowing the reliability of the measures involved is crucial because a measure that does not correlate well with itself will not correlate with something else. Ratings tend to be reasonably reliable, whereas task reliabilities are more variable (Enkavi et al., 2019). In particular, many control tasks use difference scores to isolate control processes. For example, in the Stroop task, average response time in a control condition that does not involve interference from a prepotent response is usually subtracted from time in the incongruent condition to remove variation attributable to simple speed. Although such differences improve the theoretical interpretability of the resulting measures, the reliability of difference scores can vary (Hedge et al., 2018): Even if a task has a large number of trials and generates strong experimental effects, reliability will be poor if there are not strong individual differences in the magnitude of those effects (e.g., if everyone shows a similarly sized effect).

Another issue with control tasks is the “task impurity” problem. Because control must be measured in the context of the lower-level processes that are being controlled, task performance can reflect these non-control processes (e.g., in the Stroop task, color perception, word reading, vocal speed). Task impurity is thought to be a major reason that control tasks generally show low correlations with each other (Friedman & Miyake, 2017), even when they are reasonably reliable. One solution to the task impurity problem is to administer multiple tasks that tap the same control process but differ in their lower-level requirements (e.g., the Stroop, antisaccade, and stop-signal tasks to tap response inhibition) and extract their common variance with latent variable (factor) analysis. Because latent variables capture only variance that correlates across tasks, they are purer measures and remove random measurement error (unreliability). Indeed, latent variables for task-based

measures of control show high test-retest reliability ($r_s=.86-.97$), even over intervals of 5–6 years (Friedman et al., 2016; Gustavson et al., 2018).

If tasks and ratings show low correlations because of task and unreliability and impurity, then these correlations should be much higher at the level of latent variables. Surprisingly, however, they are not (Eisenberg et al., 2019; Friedman et al., 2020; Snyder et al., 2021). For example, Snyder et al. found small correlations ($r_s=.11-.21$) of a latent variable capturing performance on nine control tasks with latent variables for the Behavioral Rating Inventory of Executive Function rating scale or the Early Adolescent Temperament Questionnaire, respectively. This pattern suggests that unreliability and task impurity cannot explain the low correlations between ratings and tasks.

Ratings may also be “impure” or invalid measures of control. Individuals’ perceptions of their own behavior may be inaccurate, or they may respond to some questions in a socially desirable manner. These factors may not simply introduce noise into survey measures, which could be removed with latent variables; instead, they may introduce systematic variation into the data (i.e., variation that might be reliable and correlate with other measures). Such biases are a major concern when considering that ratings often predict outcomes better than tasks do (Eisenberg et al., 2019; Friedman et al., 2020), because these outcomes are also typically self-reported and thus could have similar biases. For example, individuals who are reluctant to admit they have poor self-control may be similarly less forthcoming about substance use problems. Conversely, people who have difficulty managing substance use may answer questions about self-control based on these experiences. Indeed, one meta-analysis found that ratings were more associated with self-reported undesired behaviors than with observed undesired behaviors (de Ridder et al., 2012), although this difference was only statistically significant for one the rating scales examined. Thus, using objective measures of behaviors (when possible) may lead to a more balanced comparison of their associations with ratings and tasks. Ratings from other informants (e.g., parents or teachers) may also help balance the comparison, but they have their own limitations, as informants cannot know what is going on in another individual’s mind.

In addition to unreliability and task impurity, another important consideration when interpreting the low correlation between ratings and tasks is to what extent they are associated with outcomes of interest, which speaks to their criterion validity. If these measures validly assess individual differences in control, they should predict behaviors that reflect lack of control.

Ratings generally show large associations ($R^2=.03-.29$) with outcomes of interest, such as substance use and psychopathology (Eisenberg et al., 2019; Friedman et al., 2020). However, as noted earlier, some of these associations could be due to the similar self-reported mode of measurement for control ratings and behavior problems. In contrast, tasks show more variable prediction of real-world control problems. Some studies found no associations (e.g., with measures of substance use; Eisenberg et al., 2019, even at the latent variable level. Others have found significant relationships of task latent factors with substance use and psychopathology (Friedman et al., 2020; Gustavson et al., 2017), although those associations are generally smaller ($R^2=.00-.10$) than associations with ratings. Such patterns have been

interpreted as evidence that scales are better assessments of control deficits than are tasks (Barkley & Fischer, 2011; Eisenberg et al., 2019). However, meta-analytic reviews suggest that tasks are associated with a range of mental health conditions (Snyder et al., 2015). Moreover, some outcomes, namely academic behavior (Malanchini et al., 2019; Soto et al., 2020) and income/life milestones (Eisenberg et al., 2019), seem to show stronger relationships with task measures, compared to the associations of mental health outcomes with tasks.

Regardless of the relative magnitudes of their associations with behavioral outcomes, an important question is whether ratings and tasks are associated with the *same* variance in these outcomes. If ratings and tasks tap a common control ability, then they should explain overlapping variance in behavior. However, several studies that have used both ratings and tasks to predict behavior have demonstrated that they account for independent variance in outcomes (Ellingson et al., 2019; Friedman et al., 2020; Kamradt et al., 2014; Malanchini et al., 2019; Sharma et al., 2014). For example, in two independent samples, Friedman et al. (2020) found that self-reported impulsivity and task-based executive function latent variables were both significantly associated with an externalizing psychopathology latent variable in a multiple regression structural equation model that controlled for their overlap. Such patterns suggest that ratings and tasks may predict these outcomes for different reasons.

Do tasks and ratings show low correlations because they measure different constructs?

The persistently low correlations between tasks and ratings, even when measured reliably, as well as their independent associations with behavioral outcomes, suggest that they measure separable constructs. What then are these constructs? It may be that ratings and/or tasks measure something other than control; it is also possible that both ratings and tasks are valid measures of control but measure different aspects of control. Here we briefly discuss six main dimensions that distinguish them (see Fig. 2 and Dang et al., 2020 and Wennerhold & Friese, 2020 for more discussion).

1. Cognition vs. Metacognition.

As mentioned earlier, ratings are subjective measures that require the rater to have some insight into the ratee's (usually one's self) behavior and abilities, whereas tasks are more objective measures that do not require the participant to be aware of how they are doing. Thus, ratings may in part reflect differences in metacognitive awareness, while tasks assess only cognitive control.

2. Emotional Context.

Rating scales often ask whether individuals can exert control in emotional situations (e.g., impulsive urgency). Such emotional control may differ from attentional control, which is typically assessed in emotionally neutral situations. Moreover, because ratings are based on real-world experiences, they more likely reflect situations in which failures of control have negative consequences. Therefore, they may involve additional emotional investment

compared to tasks (where compensation rarely depends on performance). Research with “hot” executive function tasks (those that use emotional stimuli or include rewards; Zelazo & Carlson, 2012) provides a way to examine this possibility. A recent study implementing inhibition tasks in a gaming framework (Verdejo-Garcia et al., 2021) reported somewhat larger associations between impulsivity ratings and cognitive measures ($r_s=.2-.3$), compared to prior work, possibly because they adjusted for reliability and added motivational/emotional aspects to the cognitive tasks. However, studies have not yet examined latent variable associations between “hot” executive function tasks, standard (“cool”) executive function tasks, and rating measures.

3. Time-scale of Control.

Control assessed by computerized tasks reflects operations that take place in just a few hundred milliseconds. In contrast, rating measures ask individuals about how well they can attend to tasks in everyday life, often across very long periods of time. The attentional control demands for tasks may differ from the control necessary to make progress on long-term goals such as homework assignments or projects that take hours, days, or weeks to complete. Even when individuals must exert rapid control (e.g., when they are angry), the situation does not dissipate immediately, and they may have to continue exerting control for an extended period of time.

4. Performance in typical vs. optimal conditions.

A related distinction is that rating scales tend to ask about typical performance (i.e., across a range of contexts and occasions), whereas tasks often assess performance in optimal, experimentally constrained, conditions (Toplak et al., 2013; Wennerhold & Friese, 2020). While individuals may be capable of high levels of control during experiments in the lab (particularly when tasks include incentives for good performance), they may not necessarily use it regularly in everyday life. Thus, being “good at” control may not mean that individuals typically “act self-controlled” (Grund & Carstens, 2019). Computerized tasks that are administered online are becoming more popular and may assess performance in conditions that are somewhere between typical and optimal. Specifically, online computerized assessments require individuals to exert control in experimentally constrained situations, but even with careful study design (e.g., including items to check attention, ensuring the task will function across internet browsers and computers, instructing subjects to do the experiment in a quiet setting), there may be unexpected environmental disruptions that experimenters cannot control that influence performance.

5. Clarity of goal representation, cueing, and monitoring:

Task goals are clear in the laboratory but are not always spelled out in the real world. Even if the final goal is clear (e.g., buy bread), individuals typically need to develop and execute a plan themselves. Tasks are also typically cued, allowing individuals to know exactly when they need to exert control. Everyday situations are much more variable, requiring personal reminders or self-cueing (e.g., remembering to stop at the store on the way home). Laboratory tasks also organize distracting or irrelevant information in a systematic manner that may not represent real world contexts. Finally, error signals may also be clearer or more consistent in laboratory tasks compared to the real world.

6. Control vs. Habits.

One intriguing proposal is that ratings may not reflect active control abilities at all, but rather the use of habits that are associated with lessening the need for control (de Ridder et al., 2012; Galla & Duckworth, 2015). Individuals who score high on self-control rating scales also practice more habits like eating healthy snacks, exercising regularly, and sleeping at regular times. To the extent that individuals structure their own environments so that these behaviors become habitual (Vohs & Piquero, 2021), they no longer require control; one may have less temptation to eat unhealthy snacks if one does not keep them in the house. This possibility is consistent with findings that people who report high self-control report having to exert self-control less often (Grund & Carstens, 2019) and experience fewer occasions of temptation (Hofmann et al., 2012). If ratings reflect habits rather than abilities, then their lack of correlation with tasks is unsurprising.

These potential explanations for the incongruity between tasks and ratings are not mutually exclusive, and future research is needed to evaluate their relative importance. We hope that this brief discussion inspires studies that systematically manipulate these dimensions in both task and rating paradigms. For example, to manipulate the dimension of “clarity of goal representations” in ratings measures, researchers could ask subjects to set clear goals and establish implementation intentions for when control is needed (or assign particular implementation intentions), then ask questions about their success. These manipulations might improve the goal representations, monitoring, and cueing in ratings measures and/or reduce variance in goal representations across individuals in the sample. Another approach that may reduce ratings being influenced by insight (the “cognition vs. metacognition” dimension) would be to use experience sampling and passive data collection (e.g., with mobile phones or wearable devices) that can assess everyday control more directly, reducing reliance on retrospective judgments. For example, accelerometry and location data can be used to obtain measures related to exercise, sleep habits, and substance use (e.g., smoking hand movements; proximity to substance use outlets).

In addition to sampling behaviors under different conditions, we could gain more information about the conditions participants set up for themselves in their everyday life. Related to our earlier discussion of “control vs. habits,” if configuring the environment (reducing the need for control) is a significant aspect of individual differences in everyday self-control, it may be important to assess individuals’ abilities and spontaneous inclinations to do so. For example, self-regulation or emotion-regulation is enhanced by adopting strategies such as psychological distancing: e.g., in delay of gratification tasks such as the classic marshmallow task, thinking of marshmallows as clouds instead of tasty treats (Mischel & Baker, 1975). Interestingly, there is some evidence that instructed self-distancing (e.g., making decisions as Batman instead of oneself) can also improve children’s performance on cool executive function tasks (White & Carlson, 2016). Such results raise a number of questions: To what extent do individuals spontaneously adopt such strategies to regulate their behavior, and what factors affect the tendency to do so? Do individual differences in use of these strategies influence variance in self-control ratings and tasks? Do ratings correlate better with tasks that enable such reconfiguration compared to tasks that do

not enable reconfiguration? Experiments targeting such questions may yield insights into the separation between tasks and ratings.

As a final point, it may be useful to consider that the constructs tapped by some control measures may be broader than researchers often posit. Task-based research frequently focuses on “inhibition” tasks, with the assumption that self-control requires active suppression of impulses or distractions. However, everyday self-control and success may be more about creating good goals and monitoring the environment for cues about when to implement these goals and when to increase control. Indeed, self-control ratings often have reverse-scored items assessing whether individuals are organized, reliable, neat, and scheduled, which do not obviously reflect inhibitory control. Thus, other control tasks may be more useful than inhibition tasks, which may be particularly susceptible to poor reliability and low cross-task correlations (Friedman & Miyake, 2004; Hedge et al., 2018; Rey-Mermet et al., 2018). Similarly, it may be useful to re-evaluate whether such organizational items should be included in ratings scales as indicators of control or if some scales are conflating control with general goal-management abilities.

Concluding Remarks

In summary, rating-based and task-based control measures assess only slightly overlapping variance, even when both are measured reliably. A pessimistic view of this low convergence of tasks and ratings is that it is bad news for psychological science: If ratings and tasks indeed measure different constructs, then the specific cognitive mechanisms targeted by tasks cannot be used to understand individual differences in control as measured by ratings and vice versa. A more optimistic view is that their distinction presents an opportunity to better understand what each is really measuring through systematic investigation of the dimensions that may distinguish them. We have characterized these constructs as different “aspects” of control because we see similarities to other multidimensional constructs: Just as different kinds of artistic ability (e.g., painting and singing) may show small correlations but may nevertheless both be described as artistic, task and ratings measures may show small correlations but nevertheless both be described as control-related. However, future research is needed to evaluate whether it is indeed accurate to characterize both these constructs as aspects of a more general control ability. If ratings and tasks do capture different aspects of control, then they may be used to supplement one another, increasing prediction and insight. Evidence for such incremental validity for predicting some outcomes (Ellingson et al., 2019; Friedman et al., 2020; Kamradt et al., 2014; Malanchini et al., 2019; Sharma et al., 2014) suggests that there is reason to study both aspects of control. It may be useful to administer both sets of measures more routinely, as each may provide a different window into important individual differences.

Acknowledgements

We thank Samantha Freis, Srishti Nayak, and Roselinde Kaiser for helpful comments and discussion. Naomi Friedman was supported by National Institutes of Health grants DA046064, DA046413, DA051018, DA042742, MH117131, HD078532, and AG046938. Daniel Gustavson was supported by the National Institutes of Health grant R03 AG065643.

References

- Barkley RA, & Fischer M (2011). Predicting impairment in major life activities and occupational functioning in hyperactive children as adults: Self-reported executive function (EF) deficits versus EF tests. *Developmental Neuropsychology*, 36(2), 137–161. 10.1080/87565641.2010.549877 [PubMed: 21347918]
- Cyders MA, & Coskunpinar A (2011). Measurement of constructs using self-report and behavioral lab tasks: Is there overlap in nomothetic span and construct representation for impulsivity? *Clinical Psychology Review*, 31(6), 965–982. 10.1016/j.cpr.2011.06.001 [PubMed: 21733491]
- Dang J, King KM, & Inzlicht M (2020). Why are self-report and behavioral measures weakly correlated? *Trends in Cognitive Sciences*, 24(4), 267–269. 10.1016/j.tics.2020.01.007 [PubMed: 32160564]
- de Ridder DTD, Lensvelt-Mulders G, Finkenauer C, Stok FM, & Baumeister RF (2012). Taking stock of self-control: A meta-analysis of how trait self-control relates to a wide range of behaviors. *Personality and Social Psychology Review*, 16(1), 76–99. 10.1177/1088868311418749 [PubMed: 21878607]
- Diamond A (2013). Executive functions. *Annual Review of Psychology*, 64(1), 135–168. 10.1146/annurev-psych-113011-143750
- Duckworth AL, & Kern ML (2011). A meta-analysis of the convergent validity of self-control measures. *Journal of Research in Personality*, 45(3), 259–268. 10.1016/j.jrp.2011.02.004 [PubMed: 21643479]
- Eisenberg IW, Bissett PG, Enkavi AZ, Li J, MacKinnon DP, Marsch LA, & Poldrack RA (2019). Uncovering the structure of self-regulation through data-driven ontology discovery. *Nature Communications*, 10(1), 2319. 10.1038/s41467-019-10301-1 Found that behavioral tasks are less reliable than survey ratings when collected online. Although factor scores for tasks increase test-retest reliability, only ratings strongly predicted most substance use and mental health outcomes.
- Ellingson JM, Corley R, Hewitt JK, & Friedman NP (2019). A prospective study of alcohol involvement and the dual-systems model of adolescent risk-taking during late adolescence and emerging adulthood. *Addiction*, 114(4), 653–661. 10.1111/add.14489 [PubMed: 30398695]
- Enkavi AZ, Eisenberg IW, Bissett PG, Mazza GL, MacKinnon DP, Marsch LA, & Poldrack RA (2019). Large-scale analysis of test–retest reliabilities of self-regulation measures. *Proceedings of the National Academy of Sciences*, 116(12), 5472–5477. 10.1073/pnas.1818430116
- Friedman NP, Hatoum AS, Gustavson DE, Corley RP, Hewitt JK, & Young SE (2020). Executive functions and impulsivity are genetically distinct and independently predict psychopathology: Results from two adult twin studies. *Clinical Psychological Science*, 8(3), 519–538. 10.1177/2167702619898814 [PubMed: 33758683] Found that impulsivity ratings and executive function task latent variables based on in-person measures show low correlations and inconsistent genetic correlations and independently predict psychopathology in two adult twin samples.
- Friedman NP, & Miyake A (2004). The relations among inhibition and interference control functions: A latent-variable analysis. *Journal of Experimental Psychology: General*, 133, 101–135. [PubMed: 14979754]
- Friedman NP, & Miyake A (2017). Unity and diversity of executive functions: Individual differences as a window on cognitive structure. *Cortex*, 86, 186–204. 10.1016/j.cortex.2016.04.023 [PubMed: 27251123]
- Friedman NP, Miyake A, Altamirano LJ, Corley RP, Young SE, Rhea SA, & Hewitt JK (2016). Stability and change in executive function abilities from late adolescence to early adulthood: A longitudinal twin study. *Developmental Psychology*, 52(2), 326–340. 10.1037/dev0000075 [PubMed: 26619323]
- Galla BM, & Duckworth AL (2015). More than resisting temptation: Beneficial habits mediate the relationship between self-control and positive life outcomes. *Journal of Personality and Social Psychology*, 109(3), 508–525. 10.1037/pspp0000026 [PubMed: 25643222] Across six studies, found that the links between rating measures of control and positive outcomes (e.g., related to health and academic achievement) are explained by the use of beneficial habits.

- Grund A, & Carstens C-A (2019). Self-control motivationally reconsidered: “Acting” self-controlled is different to “being good” at self-control. *Motivation and Emotion*, 43(1), 63–81. 10.1007/s11031-018-9721-3
- Gustavson DE, Panizzon MS, Elman JA, Franz CE, Reynolds CA, Jacobson KC, Friedman NP, Xian H, Toomey R, Lyons MJ, & Kremen WS (2018). Stability of genetic and environmental influences on executive functions in midlife. *Psychology and Aging*, 33(2), 219–231. 10.1037/pag0000230 [PubMed: 29658745]
- Gustavson DE, Stallings MC, Corley RP, Miyake A, Hewitt JK, & Friedman NP (2017). Executive functions and substance use: Relations in late adolescence and early adulthood. *Journal of Abnormal Psychology*, 126(2), 257–270. 10.1037/abn0000250 [PubMed: 28045282]
- 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. 10.3758/s13428-017-0935-1 [PubMed: 28726177]
- Hofmann W, Baumeister RF, Förster G, & Vohs KD (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology*, 102(6), 1318–1335. 10.1037/a0026545 [PubMed: 22149456]
- Kamradt JM, Ullsperger JM, & Nikolas MA (2014). Executive function assessment and adult attention-deficit/hyperactivity disorder: Tasks versus ratings on the Barkley Deficits in Executive Functioning Scale. *Psychological Assessment*, 26(4), 1095–1105. 10.1037/pas0000006 [PubMed: 24885846]
- Malanchini M, Engelhardt LE, Grotzinger AD, Harden KP, & Tucker-Drob EM (2019). “Same but different”: Associations between multiple aspects of self-regulation, cognition, and academic abilities. *Journal of Personality and Social Psychology*, 117(6), 1164–1188. 10.1037/pspp0000224 [PubMed: 30550329] Found that a Common Executive Function task latent variable based on in-person measures and ratings related to self-regulation and achievement (impulse control, openness, and conscientiousness) show modest overlap and predict independent variance in academic performance in a twin sample of children (3rd through 8th grade).
- Mischel W, & Baker N (1975). Cognitive appraisals and transformations in delay behavior. *Journal of Personality and Social Psychology*, 31(2), 254–261. 10.1037/h0076272
- Rey-Mermet A, Gade M, Link to external site, this link will open in a new window, & Oberauer K (2018). Should we stop thinking about inhibition? Searching for individual and age differences in inhibition ability. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 44(4), 501–526. 10.1037/xlm0000450 [PubMed: 28956944]
- Sharma L, Markon KE, & Clark LA (2014). Toward a theory of distinct types of “impulsive” behaviors: A meta-analysis of self-report and behavioral measures. *Psychological Bulletin*, 140(2), 374–408. 10.1037/a0034418 [PubMed: 24099400]
- Snyder HR, Friedman NP, & Hankin BL (2021). Associations between task performance and self-report measures of cognitive control: Shared versus distinct abilities. *Assessment*, 28(4), 1080–1096. 10.1177/1073191120965694 [PubMed: 33084353]
- Snyder HR, Miyake A, & Hankin BL (2015). Advancing understanding of executive function impairments and psychopathology: Bridging the gap between clinical and cognitive approaches. *Frontiers in Psychology*, 6, 328. 10.3389/fpsyg.2015.00328 [PubMed: 25859234]
- Soto EF, Kofler MJ, Singh LJ, Wells EL, Irwin LN, Groves NB, & Miller CE (2020). Executive functioning rating scales: Ecologically valid or construct invalid? *Neuropsychology*, 34(6), 605–619. 10.1037/neu0000681 [PubMed: 32730048]
- Toplak ME, West RF, & Stanovich KE (2013). Practitioner Review: Do performance-based measures and ratings of executive function assess the same construct?: Performance-based and rating measures of EF. *Journal of Child Psychology and Psychiatry*, 54(2), 131–143. 10.1111/jcpp.12001 [PubMed: 23057693]
- Verdejo-Garcia A, Tiego J, Kakoschke N, Moskovsky N, Voigt K, Anderson A, Koutoulogenis J, Lubman DI, & Bellgrove MA (2021). A unified online test battery for cognitive impulsivity reveals relationships with real-world impulsive behaviours. *Nature Human Behaviour*. 10.1038/s41562-021-01127-3

- Vohs KD, & Piquero AR (2021). Self-control at 220 miles per hour: Steering and braking to achieve optimal outcomes during adolescence. *Current Directions in Psychological Science*, 30(4), 351–357. 10.1177/09637214211014509
- Wennerhold L, & Friese M (2020). Why self-report measures of self-control and inhibition tasks do not substantially correlate. *Collabra: Psychology*, 6(1), 9. 10.1525/collabra.276
- White RE, & Carlson SM (2016). What would Batman do? Self-distancing improves executive function in young children. *Developmental Science*, 19(3), 419–426. 10.1111/desc.12314 [PubMed: 25997842] Provides further discussion of the ways in which self-control ratings and inhibition tasks may assess different constructs.
- Zelazo PD, & Carlson SM (2012). Hot and cool executive function in childhood and adolescence: Development and plasticity. *Child Development Perspectives*, 6(4), 354–360. 10.1111/j.1750-8606.2012.00246.x

Domain	Specific Measures	Description	Schematic
Prepotent Response Inhibition	Antisaccade, Stop-Signal Task, Stroop ^a	Refrain from a prepotent or dominant response (e.g., stopping a reflexive eye movement to a brief cue and instead look in the opposite direction in the Antisaccade task).	
Interference Control	Flanker, Simon, Stroop ^a	Indicate the identity of a stimulus, ignoring interference from irrelevant items that have different (but not prepotent) response associations (e.g., stimuli flanking the central stimulus in the Flanker task).	
Task-Set Shifting	Task Switching Paradigms, Trail Making Test	Flexibly shift between two or more task sets (e.g., switch between categorizing stimuli based on color or shape). Subtasks are typically equally demanding and cued.	
Working Memory Updating	<i>n</i> -back, Keep Track	Monitor the environment for relevant information and continuously update the contents of working memory (e.g., remembering the past 2 spatial positions of stimuli in the 2-back test pictured here).	
Delay of Gratification, Gambling Tests, "Hot" Control Tasks	Iowa Gambling Task, Emotional Stroop	Task demands vary, but often involve learning reward/loss payoffs, or controlling responses to emotional stimuli.	

Fig. 1. Examples of task-based measures of control (e.g., executive function tasks). Task-based measures are heterogeneous, but generally assess the ability to exert control over thoughts or actions while striving to achieve one or more concrete short-term goals. Boldface type indicates the task that is illustrated in the far-right column. Condition labels are displayed above the example screenshots, and the correct answer is displayed in quotes below the screenshot. ^aSome categories can be further parsed into different subdomains and some tasks may assess multiple processes. For example, the Stroop task can be considered a measure of interference control; however, it is also used to tap prepotent response inhibition because there is typically an asymmetry in the potency of interfering stimuli (i.e., word reading is a prepotent response, whereas color naming is not).

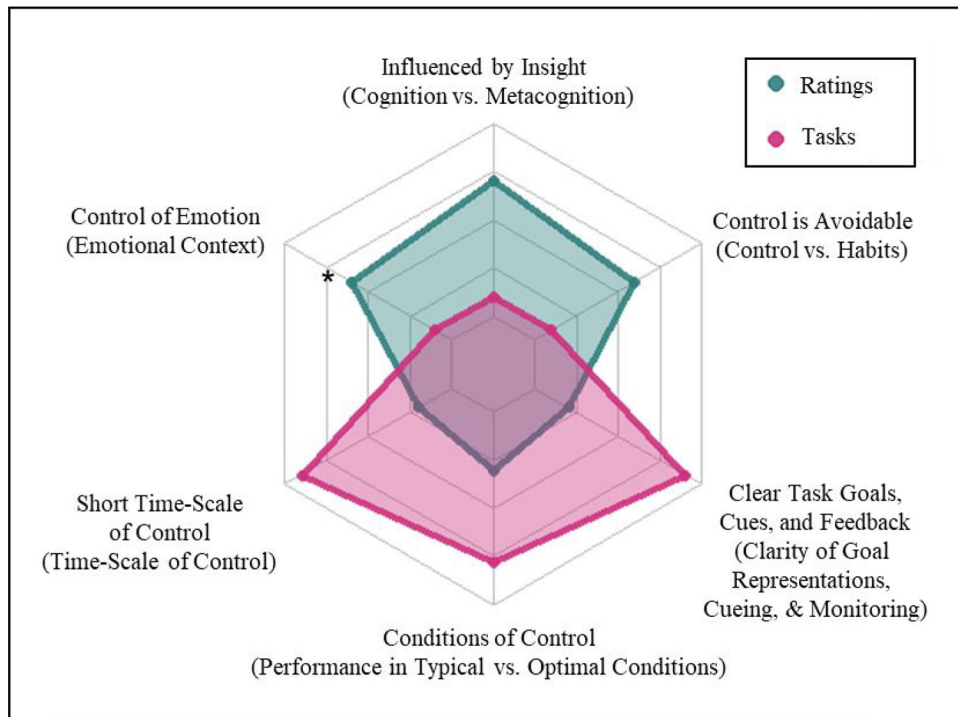


Fig. 2. Dimensions by which task-based and ratings-based measures may tap different aspects of control. For each dimension, values further away from the center reflect more consistency with the displayed label (i.e., for the top-most dimension, the higher values for rating scales mean they are more influenced by one's insight or metacognitive awareness than tasks). Task-based measures typically provide clear task goals, focus on short-term measures of control, and assess participants' performance in optimal situations. The need to exert control is typically unavoidable, and responses are not influenced by one's metacognitive awareness. Tasks typically focus on attentional control in emotionally neutral situations, though there are exceptions (i.e., hot executive function tasks); the * indicates that control of emotion is sometimes assessed by rating scales (e.g., impulsive urgency), but not always. In contrast, rating-based measures focus on long-term control and ask about how a person typically performs. The everyday control reported on in such scales are variable in terms of whether there are clear goals, cues, or feedback and whether the act of control is avoidable (e.g., by adjusting habits or the environment). They may also be influenced by metacognitive beliefs (i.e., insight into one's own abilities).

Table 1:

Examples of Rating Scales that Assess Control

Domain	Specific Measures	Description	Example Items
Executive Function	Behavior Rating Inventory of Executive Function (BRIEF)	Attentional control and organizational abilities. Subscales include inhibition, shifting, working memory, planning, emotional control, and organization.	Not available (scale is proprietary).
Impulsivity: Lack of Premeditation	Barratt Impulsiveness Scale (BIS) ; UPPS-P Impulsive Behavior Scale	Tendency to act without thinking or without control.	<i>I do things without thinking.</i> <i>I act "on impulse."</i>
Impulsivity: Impulsive Urgency	UPPS-P Impulsive Behavior Scale	Tendency to act without control when in a positive or negative mood.	<i>When I feel bad, I will often do things. I later regret in order to make myself feel better now.</i> <i>When I am very happy I tend to do things that may cause problems later in life.</i>
Resisting Temptation & Domain-Specific Control	Brief Self-Control Scale : Executive Personal Finance Scale; Three-Factor Eating Questionnaire	Ability to exert self-discipline and control impulses. Some measures include exerting control in specific situations, such as impulsive eating or spending.	<i>I wish I had more self-discipline</i> <i>I am good at resisting temptation.*</i>
Goal Management	Cognitive Failures Questionnaire , Pure Procrastination Scale	Ability to keep track of and make progress on goals in everyday life. Many items relate to prospective memory (forgetting or losing track of intended tasks) but others capture intentional but irrational delays (procrastination).	<i>Do you find you forget what you came to the shops to buy?</i> <i>I often find myself performing tasks that I had intended to do days before.</i>
Rumination or Thought Suppression	White Bear Suppression Inventory , Ruminative Responses Scale	Tendency to ruminate or inability to remove thoughts from mind.	<i>There are thoughts that keep jumping into my head.</i> <i>I have thoughts that I cannot stop.</i>
Grit	Short Grit Scale	Perseverance for long-term goals.	<i>I am diligent.</i> <i>I finish whatever I begin.</i>
Effortful Control (Temperament Dimension)	Early Adolescent Temperament Questionnaire (EATQ-R)	Assesses multiple temperament dimensions in adolescents, including activation control, attention, and inhibitory control.	<i>I put off working on projects until right before they're due.*</i> <i>I pay close attention when someone is telling me how to do something.</i> <i>It's hard for me not to open presents before I'm supposed to.*</i>
Conscientiousness (Personality Dimension)	NEO Five Factor, Big Five Inventory (BFI)	Planning, organization, thoughtfulness, and self-discipline.	<i>I am someone who...</i> <i>Tends to be disorganized.*</i> <i>Is dependable, steady.</i>

Note: "Control" and "self-control" are blanket terms for a family of functions that assess the ability to manipulate thoughts and actions. Rating-based measures of control are heterogeneous, but generally assess the ability to exert control over thoughts or actions while striving to achieve short- or long-term goals. Scales may be completed by an individual or by an informant (e.g., a parent or teacher). Some of categories listed here can be further parsed into different subdomains (e.g., impulsive urgency includes negative and positive urgency; executive function includes inhibition, shifting, working memory, etc.), and some measures may assess multiple domains. Boldface type indicates the source of the example items.

* indicates item is reverse-scored.