THESIS

VALIDATING A POINTS-BASED EFFORT EXPENDITURE FOR REWARDS TASK

Submitted by

Emily T. Sturm

Department of Psychology

In partial fulfillment of the requirements

For the Degree of Master of Science

Colorado State University

Fort Collins, Colorado

Summer 2024

Master's Committee:

Advisor: Michael L. Thomas

Carol Seger
Jaclyn Stephens
Sara Anne Tompkins

Copyright by Emily T. Sturm 2024

All Rights Reserved

ABSTRACT

VALIDATING A POINTS-BASED EFFORT EXPENDITURE FOR REWARDS TASK

Invalid performance on neurocognitive tests due to lack of effort expenditure is a concern for researchers and clinicians. Performance validity tests determine when adequate effort has been expended, but they fail to differentiate between subcomponents of effort that may be responsible for poor performance. The Effort Expenditure for Reward Task (EEfRT) is a task that provides separate measurements of reward processing and valuation constructs which could be informative as performance validity indicators. However, previous versions of the EEfRT use monetary performance-based rewards to investigate the expected value of effort, which can be problematic due to the influence of socio-economic factors and potential to systematically disadvantage participants with neurocognitive disorders. This study first aimed to examine the construct validity, specifically, the construct representation of a points-based version of the EEfRT online and in-person. The second aim of this study, which is exploratory, is to characterize patterns in embedded performance validity test performance obtained for separate neurocognitive measures as well as the EEfRT, thereby informing nomothetic span, or patterns of significant relations across measures of effort. This aim assessed whether the scores from the EEfRT indicate performance validity in other domains. Online participants (n = 342) from Prolific.com for the online sample and in-person participants (n = 27) were recruited via advertisements. Participants completed a battery including the EEfRT along with three working memory tasks, two executive functioning tasks, and one reward learning task. Results of regression analyses showed that, as hypothesized, both online and in-person participants chose hard tasks significantly more often at higher reward levels and at higher probability levels. However, contrary to expectations, a significant interaction between reward and group showed that points were more motivating in the online setting compared to in-person. Exploratory latent profile analysis revealed no clear pattern in embedded performance validity tests within the EEfRT or across other tasks. The results of this study suggest that a points-based version of the EEfRT is potentially valid for measuring effort-

based decision making, but more research is needed before it can be called an objective measure of effort in the context of validating performance on cognitive tests.

TABLE OF CONTENTS

INTRODUCTION

Effort, or willingness to expend effort, impacts nearly every aspect of cognition, but especially attention (Bruya & Tang, 2018), inhibition (Leotti & Wager, 2010), and working memory (Engström et al., 2013). Effort impacts scholastic (Jin, 2023) and diagnostic (Green, 2007; Green et al., 2001) tests and has wide-reaching, even medico-legal consequences (Moore et al., 2021). Effort-based decision making is important in a wide variety of contexts from choosing strategies to safety behaviors (Wickens, 2014). Effort deficits are tied to poorer real-world functioning, especially in the context of psychiatric and neurological disorders like schizophrenia (Tran, 2022). Malingering and insufficient effort are ever-present concerns for neuropsychologists who use standalone and embedded performance validity tests to inform their decision on whether adequate effort was expended on a test, therefore enabling interpretation of the score as a measure of ability. However, there are still disagreements about the measurement of effort, and decisions about how to use measures ultimately rely on clinical judgment (Mækelæ et al., 2023; Millis, 2008). Recent advances in the measurement of subcomponents of the effort construct may provide important information that has not yet been considered in the context of neurocognitive testing.

Defining Effort

Effort is a broad construct defined in different ways, depending on the scope and framework in which the construct is considered. Generally, effort can be described in terms of a felt application of force or ability used to perform a behavior, as mediator that improves behavioral performance until ability-level is reached the use of resources (e.g., physical or cognitive) to accomplish a behavior (Halperin & Vigotsky, 2024). Different measurement approaches are used depending on the definition of effort and the goal of the assessment.

The experience of effort expenditure is measured with self-reports that ask participants to rate perceived workload, anticipatory pleasure, motivation, or trait anhedonia (de Winter, 2014; Geaney et al., 2015; Treadway et al., 2009). Self-report measures are a quick and straightforward way to assess subjective or perceived effort. However, self-report measures are vulnerable to careless or low-effort responding (Ulitzsch et al., 2022), impacted by beliefs about self-efficacy (Hutchinson et al., 2008), and do not always show clear relationships with other measures of effort (Ohmann et al., 2022). Self-report measures of effort may be useful for measuring characteristics associated with trait motivation, but are not reasonable measures of whether sufficient effort has been expended during neurocognitive testing due to the potential incentive to respond falsely.

In the context of neurocognitive testing, effort is typically defined as a mediator between ability and task performance such that low-effort expenditure results in poor performance on a neurocognitive task that is not truly reflective of ability level, thus invalidating test results (Greher & Wodushek, 2017). In this perspective, inadequate effort is inferred based on poor test performance, typically measured as a binary variable determined by population norms (Halperin & Vigotsky, 2024). While considering effort as a mediator may be useful for explaining performance while accounting for individual capacities, this perspective lacks the nuanced definition of effort that would enable researchers to determine the reason for poor effort expenditure and possibly improve tasks to elicit higher levels of effort (Halperin & Vigotsky, 2024). Poor performance due to low effort may be caused by a number of underlying reasons that are important to consider when validating test results. For example, a participant may decide to expend little effort, which may be interpreted as malingering even though decision making is influenced by nuanced components of effort such as effort valuation. Since disorder may also impact components of effort such as effort valuation, malingering can be particularly difficult to identify (Millis, 2008). Another potential cause of the appearance of low effort impacting test scores could be that a participant actually expended a high amount of effort expended in an unhelpful way (e.g. focusing on the wrong aspect of the test) (Culbreth et al., 2018b; Halperin & Vigotsky, 2024)

Effort is often investigated with only one of two aims: to determine whether a participant was exhibiting enough effort to validate test results (Vickery et al., 2001), or to parse distinct underlying processes involved in effort-based decision-making (Assadi et al., 2009). Rarely are these two ideas of effort considered together. If effort assessments in neuropsychiatric contexts could be more informative of the underlying processes that might be at play, researchers and clinicians may be able to determine whether low effort was expended due to deficits in effort-based decision-making abilities or true malingering. When effort is defined in terms of resource allocation, effort-based decision-making approaches (i.e. opportunity-cost or neuroeconomic models) require a measure that presents a decision between effort levels related to a reward to determine the value of the felt exertion of resources (Kurzban et al., 2013). While effort is closely related to constructs like motivation, difficulty, attention, and cognitive control, effort is a distinct construct that impacts decision making that can be measured in a nuanced way (see Westbrook & Braver, 2015, for a thorough differentiation between these constructs). Within the construct of effort, effort valuation and reward processing are two important components that can be measured with tasks that take a neuroeconomic approach. Since effort is distinct from, yet necessarily related to, difficulty (Fleming et al., 2023), effort measurement paradigms offer a decision between different reward levels for engaging with varying levels of task difficulty resulting in a score that measures subjective value of effort (Salamone et al., 2018). In this framework, a typical assumption is that more difficult tasks will be avoided compared to easier tasks, unless a higher extrinsic reward is offered (Fleming et al., 2023). The addition of probability of receiving the reward furthers the utility of this approach by providing a measure of reward processing (Treadway et al., 2009). Understanding reward processing and effort valuation is important for understanding and addressing distinct underlying neural mechanisms responsible for the seemingly similar behavior of reduced effort expenditure on neurocognitive tests (Culbreth et al., 2018b; Nguyen et al., 2019). For example, participants with schizophrenia show reduced effort expenditure compared to controls in high reward and high probability conditions which has been related to both a reduced reward responsivity and deficits in cognitive control (Culbreth et al., 2018a). Knowing this, one study sought to increase reward salience by reminding participants of the monetary reward right before an effort task was performed, which increased effort expenditure (Renz et al., 2023). Therefore, if neuropsychiatric test administrators had a better understanding of which of these underlying components was responsible for reduced effort expenditure, neurocognitive tests might be adjusted to improve effort expenditure in populations with reward and valuation deficits, enabling more accurate testing of ability.

Performance Validity Tests

Despite widespread consensus that measuring effort expenditure during neurocognitive testing is essential (McGuire et al., 2019), agreed upon methods of doing so are few and far between, and agreements on how to interpret performance validity data are even rarer (McWhirter et al., 2020). In one approach, standalone measures of validity (e.g. the Test of Memory Malingering (Kulas et al., 2014)) are used to assess participants' willingness to complete very easy tasks (i.e., tasks with low performance ceilings) that are ostensibly challenging, thus allowing for the identification of inadequate effort or intentional exaggeration of cognitive deficits (Lippa, 2018). On the other hand, embedded performance validity tests use metrics derived directly from cognitive tests, such as nonresponse rate (Lance et al., 2010), extremely low accuracy (Rickards et al., 2018), or unusually fast reaction times (Stevens et al., 2016) to determine cases of insufficient effort. Embedded performance validity tests offer the convenience of enabling researchers to discard a single trial, task, or subject depending on the needs of the researcher in terms of sufficient effort. Both stand-alone and embedded performance validity tests usually have a threshold level of performance that must be reached for data to be considered valid (Bailey et al., 2018; Rickards et al., 2018). Unfortunately, thresholds change based on age (Rickards et al., 2018) and disorder (Abeare et al., 2022) and the ultimate decision of whether an acceptable amount of effort was expended is ultimately up to clinician's judgement (Slick et al., 1999).

Interpreting failure of embedded performance validity tests can be problematic. Worse performance on performance validity tests tends to be associated with psychiatric symptoms, and thus interpretative challenges are greatest for those who are in most need of accurate measurement (Bigler, 2012; McWhirter et al., 2020). Current guidance is to address this through the use of multiple embedded performance validity tests along with forensic evidence to determine whether performance is valid, invalid, or questionable (Rickards et al., 2018). While this prevents the hasty assumption that a participant with low ability was not expending good effort, it diminishes the convenience and ease of using embedded performance validity tests. Another issue with most performance validity tests is that they do not consider the importance of the impact of extrinsic or intrinsic reward on effort except in rare cases which may contribute to inequity in research, as described below (Chafetz et al., 2015). Performance validity tests measure whether sufficient effort has been expended, but do not take into account the gap between sufficient and optimal effort (Bigler, 2014). Finally, a determination of valid, invalid, or questionable does not provide the depth of information that a more nuanced measure of effort-based decision-making might offer. Instead of determining whether a certain level of effort has been reached, subconstructs of effort could be evaluated to provide a potential explanation for why performance may not have reached validity thresholds which could lead to adjustments in neurocognitive testing that may improve effort expenditure.

Effort-Construct Tasks and The Effort Expenditure for Reward Task

While effort is often viewed as a binary variable, tasks exist that are intended to measure effort in a more nuanced, quantitative way, as well as investigate the subconstructs that may be at the root of poor performance related to reduced effort expenditure (Chong et al., 2017; Keel et al., 2022; Molinaro & Collins, 2023; Nguyen et al., 2019; Waltz & Gold, 2016). Some suggest that translational effort-based decision-making tasks may even lead to improvements in drug treatments for motivational deficits (Salamone et al., 2018). One specific task, the Effort Expenditure for Rewards Task (EEfRT) (Treadway et al., 2009) is particularly notable given that it has been recommended by National Institute of Mental Health's Research Domain Criteria (*Behavioral Assessment Methods for RDoC Constructs - National Institute of Mental Health (NIMH)*, 2023), utilized with brain imaging to investigate brain activity related to effort-based decision making (Chong et al., 2017), and suggested to be an objective measure of effort (Treadway et al., 2009). The EEfRT has gained popularity for its utility in measuring factors related to effort-based decision making such as effort valuation (Zou et al., 2020), trait anhedonia (Barch et al., 2014; Treadway et al., 2009), expected value (Soder et al., 2021), anticipatory pleasure (Yang et al., 2014), and for its ability to detect unique patterns in effort expenditure related to clinical disorders (Treadway et al., 2015; Yang et al., 2014). Depression, bipolar disorder, schizophrenia, and Parkinson's disease can impact different aspects of dopaminergic systems responsible for effort-based decision making. Utilizing the EEfRT as a measure of effort expenditure in neurocognitive testing could indicate which specific neural mechanisms may be related to low effort or provide a more objective measure of global effort expenditure in healthy controls and participants with neurocognitive disorders.

The EEfRT is unique among effort-based decision-making tasks because specific subconstructs of effort can be simultaneously measured thanks to the manipulation of reward and probability during the decision-making phase of the task. The EEfRT presents participants with a choice on each trial between pressing a button less times (easy task) or more times (hard task). Additionally, the task has been robust to other kinds of effort expenditure such as cognitive effort (e.g., easy vs. hard math problems). Each easy and hard task is associated with a monetary reward, enabling researchers to understand reward valuation and reward processing based on the relationship between the percentage of hard task choices at low, medium, and high reward levels. Participants with schizophrenia show a unique relationship, making fewer high-effort choices than depressed patients and healthy controls even in high reward, high probability conditions (Wang et al., 2022). While participants with bipolar disorder expend less effort in high-reward, medium-probability conditions (Wang et al., 2022). While participants with mild depressive symptoms may have patterns of effort expenditure similar to healthy controls, severe symptoms related to anhedonia and anticipatory pleasure tend to be related to reduced effort in high-reward conditions at all probability levels (Horne et al., 2021). Thus, the manipulation of reward level and resulting impact on effort expenditure can differentiate between groups and provide a clearer understanding of the underlying neural mechanisms at play. Each easy and hard task is also associated with a probability of receiving the reward, enabling researchers to better understand the interplay between reward valuation, effort valuation, expected value, and decision-making in uncertain conditions (i.e. low probability levels) (Nguyen et al., 2019). The interaction between reward and probability is usually interpreted as expected value and can also be impacted differently by different disorders.

Utilizing the metrics from this task as measures of effort expenditure could potentially be applied as a way of measuring the specific aspects of effort-based decision processing that may be causing invalid test scores, and remedies addressing the subconstructs of effort could be created. Specifically, people with schizophrenia, major depressive disorder, and bipolar disorder have distinct patterns of response and patterns, especially pertaining to the interaction between reward and probability, of brain activity during the choice phase of the task (Wang et al., 2022). If the EEfRT could be used as a standalone performance validity test within a battery of cognitive tests, the interaction between reward and probability could provide clues to the neural systems responsible for decreased effort expenditure. However, it is not yet known if the expected value score in the EEfRT relates to measures of low effort in other cognitive tasks.

Despite the ability of this task to provide a nuanced picture of effort expenditure as a construct, the EEfRT is not typically used as a standalone measure of performance validity within neurocognitive testing batteries and when it is, total percentage of hard tasks chosen is the only metric that has been identified as an indication of low effort. One reason for this could be the lack of agreed upon validity metrics within the task itself. Few studies using this task check embedded performance validity metrics and those that do typically exclude participants who did not choose a mixture of easy and hard tasks. If the decision to perform easier tasks over harder tasks relates to measures of low-effort on other tasks, perhaps the decision to expend low effort could be better understood in opposition to the appearance of low-effort due to low ability level. No study has yet investigated whether non-response rate during the decision-making phase of the EEfRT relates to non-response rates in other tasks. If non-response rate is consistent across various types of cognitive tasks, it may be suggestive of a global low-effort expenditure while a pattern of high non-response rate on a select few cognitive measure could mean the participant has difficulties with the task itself and is not selecting to expend low-effort intentionally.

Validity of the EEfRT as a Measure of the Effort Construct

Many studies have replicated the original findings that participants tend to choose the hard task more often when the monetary reward level is increased (Culbreth et al., 2020; Geaney et al., 2015; Lopez-Gamundi & Wardle, 2018; Ohmann et al., 2018, 2022; Treadway et al., 2009). This relationship is a critical component of the task’s construct validity. Specifically, construct representation, which is the representation of a psychological construct by a response or pattern of responses across conditions in an experiment (M. E. Strauss & Smith, 2009). Since the pattern of increasing hard task choices with increasing reward and probability is altered in in populations with anhedonia and reduced trait anticipatory pleasure that also demonstrate reduced effort on neurocognitive tests, this task has been suggested to be an "objective" measure of effort that is not susceptible to the problems, described above, with embedded performance validity tests and self-reports (Barch et al., 2014; Geaney et al., 2015; Treadway et al., 2009, 2015). Low effort on cognitive tests may be due to reduced ability to evaluate effort and reward rather than simply "not trying hard enough". Higher scores on self-report measures of anhedonia have been shown to correlate with a decreased percentage of hard tasks chosen on the EEfRT (Treadway et al., 2009) such as the Chapman Social and Physical Anhedonia measure in controls (Barch et al., 2014). For participants with schizophrenia, decreased percentage of hard tasks chosen has been shown to correlate with higher scores on the Beck Depression Inventory in participants (Barch et al., 2014; Yang et al., 2014), the Snaith–Hamilton Pleasure Scale, and Temporal Experience of Pleasure Scale in participants with schizophrenia (Yang et al., 2014). The consistent relationships between self-report measures of anhedonia and scores on the EEfRT has led some to suggest that the EEfRT could be considered a useful alternative to self-report as an "objective" measure of effort (Ohmann et al., 2022; Treadway et al., 2009).

Increased probability of receiving a reward has been consistently shown to increase hard-task choices, demonstrating in-tact reward responsivity and enabling reward valuation to be understood more deeply (Culbreth et al., 2020; Geaney et al., 2015; Lopez-Gamundi & Wardle, 2018; Ohmann et al., 2018). More specifically, there seems to be an interaction between reward and probability (sometimes referred to as expected value) that drives decision making (Geaney et al., 2015; Ohmann et al., 2018, 2022), but not always (Culbreth et al., 2020). Decreased selection of hard tasks at high reward levels and increased selection of hard tasks compared to controls at low levels of probability are related to clinical high risk for schizophrenia suggesting that unique patterns in reward and probability response may be related to factors other than willingness to expend reward and may be more related to issues in dopaminergic systems responsible for decision-making (G. P. Strauss et al., 2023). The addition of probability during the decision-making phase of the EEfRT is unique compared to other effort-based decision-making tasks like the breakpoint ratio task (Keel et al., 2022), the grip effort task (Reddy et al., 2015), and the balloon analog risk task (Lejuez et al., 2002). The balloon analog risk task does include a probability component, but since the probability is not explicit, that task may be better suited for measuring effort in the context of risk taking (Lauriola et al., 2014).

One challenge to suggesting the EEfRT as an “objective” measure of effort is the lack of agreement about what constitutes valid performance within the task itself. One performance validity metric that is often checked when the EEfRT is used or modified is whether participants choose a mixture of easy and hard tasks and it has been considered noncompliance if a participant did not select a mixture of easy and hard (Geaney et al., 2015; Ohmann et al., 2018). High completion rate of trials has also been used to determine whether participants were likely fatigued or had motor-capability to perform the task (Geaney et al., 2015) although hard trials typically have a lower completion rate than easy trials (Ohmann et al., 2018). Fatigue may also play a key role in participant performance. Significant negative associations between number of hard tasks chosen and trial number have been found repeatedly and are typically interpreted as a sign of fatigue (Geaney et al., 2015; Hughes et al., 2015; Ohmann et al., 2018; Treadway et al., 2012), this is usually accounted for by including trial number as a covariate in analyses, but could be useful in the future for investigating patterns in subconstructs of effort expenditure. High nonresponse rate has also been used to invalidate results, especially if a participant made no responses for the entirety of the task (Hughes et al., 2015). While nonresponse may denote participant noncompliance, more investigation is needed to determine whether the reward payment system may be the root cause.

There are also important individual differences across diagnostic groups like schizophrenia and major depression. One analysis found that older individuals with greater cognitive deficits tend to utilize the reward and probability information less when selecting hard or easy tasks (Saperia et al., 2023). This could mean that older and impaired individuals are not as sensitive to reward and probability information during neurocognitive testing. Since it has been shown that stressing the importance of a reward can improve performance (Renz et al., 2022), perhaps interventions could be designed to ensure that participants with reduced ability to integrate reward and probability information are given adequate assistance in understanding the importance of expending effort in neurocognitive test situations or inclusion of reward and probability manipulations within neurocognitive testing could improve assessors' ability to determine whether decision-making is truly impaired or a result of malingering.

Equity and Fairness in Administration of the EEfRT

Payment and Reward

Research is guided by rules of beneficence, but paradigms that investigate reward learning often utilize performance-based payments to investigate response to manipulations of reward level. Although the amounts of these payments are guided by the rule to not cause undue induction, what is often not considered is whether it is ethical to pay participants based on performance when a task is known to elicit worse performance in a specific group or population. Additionally, the small amount of money offered as a monetary incentive may have different subjective values to participants who have lower socio-economic status and there is no guidance for when to update payment for research in the event of an unusually high rate of inflation or in countries with different currencies. In research focused on investigating the construct of effort, monetary incentives are often used. However, in some research environments (e.g., smaller schools, more rural schools, and schools with less research funding), paying participants an extra fee as a reward may not be feasible.

Monetary incentive has generally been shown to improve validity indices on cognitive tests compared to no incentive (Merritt et al., 2019), especially for those at the lowest levels of ability (Rydval & Ortmann, 2004). The amount of monetary reward offered, and even the importance placed on the reward by researchers (Renz et al., 2022) plays an important role in this association (Bijleveld et al., 2023). This is of particular interest to researchers investigating cognitive deficits and effort on cognitive testing more broadly because cognitive tests do not typically involve a monetary incentive leaving intrinsic motivation as the main driving factor in performance which may be impacted by test-taking context (Chafetz et al., 2011). In fact, it is sometimes assumed that participants are incentivized to reduce effort in the case of testing related to social security benefits (Schroeder et al., 2022). On the other hand, in the case of neurocognitive testing for athletes, there are differences in motivation between baseline and post-injury testing related to the incentive to return to play (Rabinowitz et al., 2015).

The EEfRT typically offers a maximum reward of approximately $5.00 per trial. Due to the need for many trials to conduct statistical analysis, the total amount earned would be remarkably high if participants were paid their total earnings. While monetary reward versus no reward has improved scores on other neurocognitive tests in participants with traumatic brain injury, suggesting that performance on cognitive tests meant to measure ability may be improved by adding a motivational component (Shum et al., 2004), paying a large amount for research participation is often impossible due to limited funding for research studies. Undue inducement, or paying research participants so much that they would be more likely to accept risks of a research study must also be avoided (Emanuel et al., 2016). On the other hand, some suggest that for low-risk research such as easy neurocognitive tests, there should not be reluctance to pay high amounts (Savulescu, 2001). To measure effort in a behavioral economics context, perhaps increasing payment to a more meaningful amount should be explored. However, given the increasing priority of reducing costs in neurocognitive testing, adding a reward payment does not seem feasible (Howieson, 2019).

One solution that has been used with the EEfRT in the past has been to deceive participants. When using deception to avoid paying participants for every trial, participants believe that they will either receive the full amount of their winnings or an amount from randomly selected trials. However, payment is the same amount at the end of the study for all participants regardless of performance. Ethically, deception in research requires that no other nondeceptive method is available (Boynton et al., 2013), but in this case, the only other method that has been used (and not tested for validity) was utilizing hypothetical money, creating the problems described below. Additionally, there is evidence of monetary rewards decreasing intrinsic motivation which may negatively impact other tests in the battery if a monetary effort-based decision making task was used as a performance validity measure (Murayama et al., 2010).

Even when self-reports suggest that monetary reward was of subjectively similar value to all participants, the amount and context of the monetary reward have been shown to impact effort-based decision-making. Studies using low and high rewards that differed by less than a dollar showed improved performance for high rewards compared to low rewards (Herrera et al., 2014; Zedelius et al., 2012) so changing the monetary reward even slightly or due to factors like inflation may have an impact on performance (Ostaszewski et al., 1998). One study that utilized the EEfRT found the increase in hard-task choices was greatest between low to medium reward levels with only a minor increase in number of hard tasks chosen from medium to high reward levels, suggesting that the precise amount of the reward may be an important factor in effort-based decision making (Renz et al., 2023). When researchers mentioned the importance of the monetary reward to participants before completion of the EEfRT, reward salience and number of hard task-choices increased (Renz et al., 2022). Another study found that when participants felt safe, the monetary reward had less impact on their decision-making (B. Schmidt et al., 2020), so emotional context may also impact the effect of monetary reward. Thus, the importance and impact of the monetary reward may not be the same across studies.

Another solution to the dilemma of paying participants for performance on the EEfRT has been to tell participants to imagine the money is real, but that all monetary values are "hypothetical". Despite some evidence supporting the use of hypothetical money in samples of healthy undergraduate samples in other tasks, the EEfRT has not been examined for differences in decision making related to hypothetical reward (Locey et al., 2011). This solution is especially problematic when considered for use with participants with neurocognitive disorder, who show differences in hypothetical decision-making compared to controls (Brown et al., 2013). Specifically, participants with schizophrenia show higher delay discounting for real money, compared to hypothetical money (Horan et al., 2017). A non-hypothetical money reward might provide a useful alternative to hypothetical money in this context.

Validating a points-based version of the EEfRT not only improves upon the previously described issues with monetary reward, but also contributes to the understanding of the utility of gamification in neurocognitive testing. Gamification, or the use of game like attributes has recently been investigated as a promising way to improve engagement in neurocognitive testing (Lumsden et al., 2016). Leaderboards, badges, and points are generally found to be the most common game elements used in when gamifying an existing task (Hishamuddin Abdul Rahman et al., 2018). Leaderboards, or a list of participants with the highest scores on a task that is shown to future participants to motivate them in a competitive way, is one of the least studied gamification elements. Since the EEfRT is likely to be used with participants with schizophrenia and other neuropsychiatric disorders that are associated with social amotivation or aberrant social processing, competitive game elements may not be a good choice (Chaldogeridis & Tsiatsos, 2022). Badges are collectable tokens of achievement, awarded for completing subgoals or levels within a task (Berntzen et al., 2013). However, badges may only increase motivation for performance when the behavior being reinforced is ultimately viewed as useful to the participant, and require maintenance of goals during task performance which may impact the measurement of effort-based decision making (Denny, 2013). Additionally, the EEfRT is not well-suited to the addition of badges due to the lack of sub-goals or levels inherent in the task. Rewarding performance with points has been shown to increase subjective engagement (Miranda & Palmer, 2014), and improve reaction time and error rates (Wiley et al., 2020). Points have been shown to have a similar motivating impact to small amounts of cash (Lumsden et al., 2016), and have been successfully utilized in neurocognitively diverse populations such as participants with ADHD (Chaldogeridis & Tsiatsos, 2022). Other gamification rewards include tangible rewards like small prizes, animated feedback, and positive peer response, but points appear to be the most effective of these (Lewis et al., 2016). Still, care must be taken to ensure that gamification elements do not distract participants from the task (Bekk et al., 2022). Utilizing multiple game elements and displaying point totals during active task engagement have been shown to increase participant distraction, so displaying point totals at the end of every trial during a feedback phase is the recommendation followed in the current study (Bekk et al., 2022; Lumsden et al., 2016).

Context/Setting of Administration: Online vs. In-person

Another way to improve equity in research is by making research participation accessible to more diverse sample populations with online administration. Digital equity initiatives are working to improve access to technology and internet access in underserved populations, supporting the diversity in online samples (Banerjee, 2020; National Telecommunications and Information Administration, 2023). During the COVID-19 pandemic, remote neurocognitive testing gained in popularity, enhancing the need for validated online paradigms (Webb et al., 2022). Currently, barriers to in-person testing such as transportation costs, physical disability, and cost of administration of in-person tests make online testing a viable way to reach a diverse population (Requena-Komuro et al., 2022; Van Patten, 2021). Specifically, the online recruitment platform, Prolific, enables researchers to choose participants based on demographic characteristics to ensure a representative or diverse sample (Tomczak et al., 2023). Although Prolific does enable participants to message the individuals running the study, it would be much less likely for an online participant to ask clarifying questions if they are unsure of what they are supposed to do for the task and there may be an impact of the lack of social pressure from a researcher's presence. On the other hand, in-person testing creates a greater opportunity for tester-induced bias, such as one group of participants systematically receiving more help understanding the tasks. Yet, use of the task online, especially a points-based version, may address the equity issues discussed above.

Embedded Performance Validity Metrics of the Points-Based EEfRT Online and In-Person

A more nuanced measure of effort in the context of neurocognitive testing is needed, and the EEfRT is a good candidate due to the multiple measures that can be extracted from a single task to illuminate the underlying subcomponents of effort-based decision-making that may be at the root of poor performance. Additionally, a points-based version of the EEfRT should be explored due to the ethical and practical concerns regarding paying for performance, increased availability, and relevance to gamification. Many researchers use the monetary version of the EEfRT to investigate relationships between willingness to expend reward and symptoms of anhedonia and amotivation in healthy controls and participants with disorders like schizophrenia (Culbreth et al., 2023).

One aspect that is important to consider whether online test setting impacts whether participants expend adequate effort on the EEfRT before testing for construct validity. Currently, there is no consensus on the best embedded performance validity tests to use with the EEfRT. Some previous studies have used completion rate (Byrne et al., 2023), while others have assessed validity by determining whether participants choose a mixture of easy and hard tasks (Treadway et al., 2009). Since current guidelines suggest that multiple performance validity tests should be failed before discarding data (Lippa, 2018), multiple potential performance validity tests will be investigated. Identifying useful performance validity metrics could improve the ability of future researchers to eliminate data which obscures the relationships between effort-based decision-making and other factors. Despite some evidence that unsupervised online cognitive testing can produce valid results (Feenstra et al., 2018; Wesnes et al., 2017), some studies suggest worse performance in online testing scenarios (Belleville et al., 2023; Morrissey et al., 2023). It is hypothesized that online participants will have higher rates of failed embedded performance validity metrics; specifically, non-responses, responses that occurred faster than a person would have been able to read the reward and probabilities offered, non-completion of button pressing trials, fatigue, and selection of easy tasks.

Validity of Points as Reward for Manipulating Decision-Making

This study aims to investigate whether points are an effective replacement for monetary reward online and in-person by relating reward to percentage of hard tasks chosen online and in-person. As previously reviewed, prior findings with the monetary version of the task show that participants choose hard tasks more often with increasing reward. Showing that increasing the points functions to increase effort expenditure is critical to establishing the construct representation validity of the points-based EEfRT. The importance of this task comes from the ability to break down the relationships between reward processing (reward manipulation), risk aversion (probability manipulation), and effort valuation (expected value). Thus, if results show a main effect of reward, then increased points successfully motivated the participants to increase effort expenditure, indicating that the points-based reward was sufficient to manipulate effort expenditure. If the results show a main effect of probability, sensitivity to increased risk in low-probability scenarios would be supported. Additionally, if there is an interaction between reward and probability such that percentage of hard tasks is greatest at highest expected value, this would suggest the task manipulations work to increase expected value at the highest levels of reward and probability and may be a good candidate for investigating effort subconstructs as the monetary version has done in the past. It is hypothesized that participants will be more likely to choose hard tasks at higher levels of reward and probability and that online test setting will attenuate this relationship.

Profiles of Effort Expenditure in Neurocognitive Testing

The EEfRT has been suggested to be an "objective" measure of the effort construct, but the EEfRT has only rarely been used as a standalone performance validity test, despite the aim of embedded performance validity tests to measure adequate effort expenditure. This could be, in part, because patterns between the effort-based decision-making measures from the EEfRT and performance validity tests on other cognitive tests have not been explored. The second, exploratory, aim of this study, will be to investigate patterns in embedded performance validity tests across six cognitive tasks, including the EEfRT, all administered online. Nomothetic span is the pattern of significant relations between measures of the same or different constructs (Whitely, 1983). Exploring the relationships between traditional embedded performance validity tests like non-response and scores from the EEfRT enables investigation into the nomothetic span of this novel version of the EEfRT. In addition to the EEfRT, three working memory tasks (N-back, Sternberg), two executive functioning tasks (Continuous Pointing Task (CPT), Flanker), and one reward learning (Probabilistic Learning Task (PLT)) were administered. Investigating patterns of effort expenditure across different cognitive domains may reveal whether this physical effort task measures the same effort construct measured by embedded performance validity tests. If EEfRT scores are indicative of a global level of effort on all online cognitive tests the EEfRT may be useful in place of multiple embedded performance validity tests within each task. Nomothetic span, or pattern of significant relationships between measures of the same or different constructs will be explored within this aim by utilizing non-response and proportion of hard tasks chosen in the EEfRT to categorize participants into different profiles of effort expenditure.

Methods

Participants

Online Participants

Online participants were pre-selected via Prolific.com based on demographic characteristics tracked by Prolific; specifically, age, sex, race, and income. After reading the consent, participants used a desktop or laptop computer (not phone or tablet) to complete the study. Once participants agreed to use a desktop or laptop computer, they provided consent and then answered a survey, hosted by Qualtrics.com, including detailed demographic information questions. After the survey, participants completed all six cognitive tasks in a randomized order. Online participants were included if they completed all six cognitive tests and the survey. All participants were paid $8.00 per hour. The median completion time for all tasks and the survey was approximately 2 hours. Online participants were included if they passed an attention check during the Qualtrics survey and completed all six cognitive tasks.

In-Person Participants.

Flyers and advertisements in print in the Front Range area of Colorado were used to recruit in-person participants as part of a larger study investigating neurocognitive measurement in participants with schizophrenia and healthy controls. Only healthy controls were included in the present analyses. In-person participants first completed a phone screen and then a four-hour in-person visit that included a structured clinical interview conducted by trained research assistants (Sheehan et al., 1997), detailed demographic survey, and completion of the battery of six cognitive tasks. Participants who were eligible for a magnetic resonance imaging scan were invited back for a second visit, but these data are not analyzed here. While only one version of the EEfRT was administered, there were two versions of the other six cognitive tasks, one of which was adaptive and will not be analyzed here. All tasks were completed on a laptop computer. Administration order of the six tasks was randomized. The two versions were randomized between the first and second visits across participants. Participants were included in this study if they were 18-70 years old, had adequate hearing and eyesight, were fluent in English, and were able to perform the cognitive and imaging tasks required. Participants were excluded if they had ever met diagnostic criteria for a schizophrenia spectrum disorder, bipolar disorder, or cluster A personality disorder, had a first-degree family member who has ever met criteria for a psychotic disorder, or were taking psychoactive medication. Participants with schizophrenia spectrum disorders also participated in the study, but these data are not included in these analyses. Finally, participants were required to pass the University of California, San Diego Brief Assessment of Capacity to Consent (UBACC).

Effort Expenditure for Reward Task

A points-based version of the EEfRT used in this study was created in Psychopy3 (see **Figure 1**). The task began with two trials in which participants were presented with a large white circle on a black background and instructed to press the spacebar as fast as possible, first with their dominant index finger for 7 seconds and then with their nondominant pinky for 14 seconds. Then, participants received instructions explaining that on each trial, they would have to choose between an easy or hard task and were given two sample trials to demonstrate the difference in task difficulty. Progress in the task was shown visually to the participants by a small red circle that expanded with each button press until it filled the area of a white circle on the black background (thus giving the impression of “popping the balloon”). Identical to the monetary version, easy trials required the participant to press the spacebar 30 times in seven seconds with their dominant index finger and were always worth the same amount of reward. In previously used monetary versions, $1.00 is typically offered for all easy trials, while in the points-based version used here offered 100 points. Hard trials required participants to press the spacebar 100 times in 21 seconds using their nondominant pinky finger, and the number of possible points varied with difficulty. In the monetary version of the task, $1.50 - $5.00 is typically offered. We offered 150 to 800 points since there was no monetary limitation to this number and the maximum possible point total was not such a large number that it could be difficult to perceive on the screen. Participants were told that they would have a certain probability of winning points on each trial if they successfully completed the trial. Regardless of whether the participant chose easy or hard, the probability of winning points on a given trial was the same (and was indicated at the bottom of the decision screen). If they did not choose a difficulty level after 7s, a difficulty level was selected for them at random, and these trials were considered a nonresponse. After each trial, participants received feedback indicating whether they popped the balloon, the number of points they won, and their updated total point score, which was presented on screen for 3 seconds after each trial. Participants could then choose to repeat the practice session or move on.

During the test segment, point values were identical to the practice. There were three trials at each point value, each with a different probability of reward (12%, 50%, or 88%). All trials were randomized. Participants received feedback on whether they popped the balloon, how many points they won, and their updated total point score. The test phase lasted approximately 10 minutes. Since the hard trials were longer than the easy trials, total time on task depended on how many hard trials were chosen, but participants were not alerted to this fact.

Previously used embedded performance validity metrics for the EEfRT were compared between online and in-person samples in order to test the hypothesis that online participants will expend less effort on cognitive testing compared to in-person participants. The embedded performance validity measures included non-response rate during the decision phase of the task, responses made in less time than would be needed to perceive the stimuli, non-completion of the button pressing phase of the task, and selecting a mixture of easy and hard tasks, and fatigue resulting in less hard-task choices in later trials. Fast responses were calculated as responses during the decision-making phase within the first 600ms of the decision-making stimuli appearing on the screen. This time limit was selected because the decision-making screen contains three vital pieces of information that should be viewed and considered before making a response, the reward for the easy task, the reward for the hard task, and the probability of receiving the award. Assuming a reading rate of 200ms per piece of information, a response of less than 600ms would suggest that the participant was not evaluating the decision, but relying on another strategy to determine their response. Non-completion of the button pressing phase of the task was calculated as any trial in which the participant failed to pop the balloon.

The EEfRT non-response score was obtained by first creating a variable where one represented a response and 0 represented a non-response. Then, means were calculated for each participant and transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. Distribution of the z-scores was visually inspected for normality and the skewness was -5.35, indicating the data were highly skewed to the right.

A second EEfRT responsivity score was also created. Since the EEfRT includes an effort-based decision-making component, the decision to expend more effort by choosing the hard task could also be considered an indicator of "objective" effort. Therefore, a variable was created where one represented a hard-task choice and 0 represented an easy-task choice. The means of this variable were calculated for each participant by dividing the sum by the total number of choices (non-responses not included in total). Then, means were transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. Distribution of the z-scores was visually inspected for normality and the skewness was 0.34, indicating normal distribution.

Cognitive Tests

The six cognitive tests administered in addition to the EEfRT included the CPT, Flanker, N-Back, PLT, and Sternberg task. This particular set of cognitive tests was chosen due to the National Institute of Mental Health Research Domain Criteria listing each of them as important measures of cognitive constructs that may be informative in the future for better understanding specific neurocognitive mechanisms related to neuropsychiatric disorder (National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (RDoC), 2016). Additionally, the tasks provide representation for both the most common and newer paradigms that are quickly gaining popularity in neurocognitive research. Better understanding the patterns of performance validity among these tasks is critical for preventing and understanding the confounding impact of cognitive effort-based decision-making on cognitive task scores.

The CPT is a measure of sustained attention and cognitive control, first developed in 1956 and later updated to a useful translational version in 2009 on which the version used in this study is based

(Roebuck et al., 2016; Rosvold et al., 1956; Young et al., 2009). Participants saw a white circle appear above one of four white lines on a black background. When a circle appeared above a line, the participants pressed a corresponding key on their desktop or laptop keyboard. A backward mask, consisting of four white squares, appeared after the circle(s) to ensure that participants could not determine the circle location from afterimages; this also allowed for manipulation of task difficulty. Stimulus durations (0.25, 0.15, and 0.1s) and mask durations (0.75, 0.85, and 0.9s) varied, with shorter stimulus duration paired with longer mask periods. Inter-trial intervals varied in duration (0.5, 1.0, and 1.5s) after each trial. Participants completed a practice block with feedback after each practice trial for 2 seconds, indicating response accuracy (correct or incorrect). After each block of trials, there was a brief blank screen lasting 3 seconds. The testing segment lasted 5 minutes and one second and included 16 test blocks of eight trials each.

The CPT non-response score was obtained by first creating a variable where one represented a response and 0 represented a non-response. Then, means were calculated for each participant and transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. The distribution of the z-scores was visually inspected for normality and the skewness was 0.78, indicating a moderate skew to the right.

A second CPT responsivity score was also created. Half of the trials in this task require a non-response as a correct answer. Therefore, over-responding could also be an indicator of low effort. A responsivity score was calculated by subtracting .5 from the mean response rate, calculating the absolute value of that difference, and flipping the sign so that a higher score represents a response pattern closest to highest accuracy, interpreted to be highest effort in this context. This variable was also visually checked for normality and was found to be severely skewed to the left with a skewness of 1.70.

The Flanker was developed in 1974 and updated more recently for use with the highly validated National Institutes of Health (NIH) Toolbox Cognitive Battery on which the version used in this study is based (Eriksen & Eriksen, 1974; Zelazo et al., 2014). Participants saw a row of arrows and used the arrow keys on their keyboard to indicate whether the arrow in the center of the screen pointed right or left.

Participants first completed a 2-minute practice segment, which included 1 block of eight trials. For each trial, participants were presented with a black screen containing a horizontal row of seven arrows that either matched the direction of the center arrow or did not match the direction of the center arrow. During the practice block, stimuli were presented for one second, during which time the participant could respond to the target's orientation. Feedback indicating response accuracy was presented for one second, and a fixation cross appeared for the two seconds in between the trials. The test phase lasted approximately five minutes and included 20 blocks containing six trials each. During the testing blocks, stimuli durations were only 0.25. Inter-trial interval duration varied between 0.5, 1.0, and 1.5s. Blocks were separated by three seconds.

The Flanker non-response score was obtained by first creating a variable where one represented a response and 0 represented a non-response. Then, means were calculated for each participant and transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. Distribution of the z-scores was visually inspected for normality and the skewness was -5.33, indicating the data were highly skewed to the right.

The N-Back is a working memory task that was first developed in 1958 and updated in 2017 to be valid for use with participants with schizophrenia on which the version used in this study is based (Kirchner, 1958; Thomas et al., 2017). Participants were presented with white 3-letter pseudowords on a black background and used the spacebar to indicate if they had seen the current word a certain number of words ago (either 1, 2, 3, or 4 words back, depending on the block). Participants began with a practice trial during which they were instructed on the definition of the n-bac conditions then, 3-letter pseudowords (e.g., CAC, NOX, GUX, VIV) were presented on the screen one at a time. For each condition, participants had to press the spacebar when a pseudoword was repeated after 1, 2, 3 or 4 words, and they practiced each condition with feedback on the accuracy of their performance. The practice phase included ten items for each condition, with each item being displayed for 2.5s followed by a blank screen for 0.5s. Participants could respond at any point, while the word was displayed or during the blank screen. Feedback was displayed for one second before moving on to the next trial. Participants then moved on to the test phase, which included no feedback and consisted of 25 items per condition, the condition blocks were randomized.

The N-Back non-response score was obtained by first creating a variable where one represented a response and 0 represented a non-response. Then, means were calculated for each participant and transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. Distribution of the z-scores was visually inspected for normality and the skewness was 0.56, indicating the data were moderately skewed to the right.

A second N-Back responsivity score was also created. Only 30% of the trials in this task require a response as a correct answer. Therefore, over-responding could also be an indicator of low effort. A responsivity score was calculated by subtracting .3 from the mean response rate, calculating the absolute value of that difference, and flipping the sign so that a higher score represents a response pattern closest to highest accuracy, interpreted to be highest effort in this context. This variable was also visually checked for normality and was found to be severely skewed to the left with a skewness of -2.44.

The PLT is a measure of probabilistic feedback processing and was first developed in 2009 then updated in 2018 for use with participants with schizophrenia on which the version in the current study is based (Bismark et al., 2018; van den Bos et al., 2012). Participants saw a pair of white geometric shapes (e.g., a square and a circle) on a black background and were instructed to choose the shape that was "Correct" (i.e., rewarded with positive feedback) more often using the left and right arrow keys on their keyboard. Shapes were displayed for three seconds, during which a response had to be made. Participants received feedback after each trial indicating whether the shape they selected was "Correct" or "Incorrect." For the practice, choosing the rectangle produced "Correct" on ~80% of the trials (11 of 14 practice trials), while choosing the plaque produced "Correct" on only ~20% of the trials (3 of 14 practice trials). After the feedback was presented for one second, there was a one second inter-trial interval. Participants could choose to repeat the practice or move on. The test segment consisted of 20 trials for each of 4 blocks, each block with different pairs of shapes. Within each pair, the probabilities were 60/40, 70/30, 80/20, 90/10, with 60/40 representing very disruptive feedback and 90/10 representing very consistent feedback. The test segment lasted approximately 7 minutes.

The PLT non-response score was obtained by first creating a variable where one represented a response and 0 represented a non-response. Then, means were calculated for each participant and transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. Distribution of the z-scores was visually inspected for normality and the skewness was -12.46, indicating the data were highly skewed to the left.

The Sternberg task is a working memory task that was first developed in 1966 and updated in 2012 to use simultaneous instead of sequential presentation of the stimuli, on which the version used in this study was based (Okuhata et al., 2012; Sternberg, 1966). Participants viewed a black screen with white upper-case letters displayed in a grid with four rows and four columns. After a maintenance period, the next screen presented a single letter, and participants used the left and right arrow keys on their keyboard to indicate whether the probe was a letter they had seen within the previous set of letters. Participants completed a practice block of eight trials. The number of letters to memorize varied, having 2, 4, 6, 8, 10, or 12 letters. Encoding duration and maintenance duration also varied by trial, lasting 2 or 6s. The probe was presented to participants for four seconds, followed by an inter-trial interval averaging 2s. Response feedback was provided after each trial during the practice block, but not the test phase. The test phase lasted approximately six minutes and included 24 trials.

The Sternberg non-response score was obtained by first creating a variable where one represented a response and 0 represented a non-response. Then, means were calculated for each participant and transformed into z-scores by subtracting the group mean from the participant mean and dividing by the standard deviation. Distribution of the z-scores was visually inspected for normality and the skewness was -4.91, indicating the data were highly skewed to the left.

Data analysis

Embedded Performance Validity Measures for the EEfRT

Due to the non-normal, zero-inflated Poisson distribution of non-response, fast response, and non-completion, zero-inflated Poisson regressions were used for these variables (Loeys et al., 2012). . Zero-inflated Poisson regression separately models the number of zeros in the data set and the non-zero counts, enabling interpretation of each component separately. The zero-inflated model allows for the interpretation of estimated odds of observing excess zero in online participants. The count model allows for the interpretation of estimated odds of observing an increase in the number of failed embedded performance validity tests in the group of participants that failed had at least one trial fail on the embedded performance validity tests. Multiple linear regression was used for normally distributed variables (mixture of hard and easy choices and fatigue), using the lme4 package in R version 4.3.2 and zero-inflated Poisson regressions was used for zero-inflated Poisson distributed variables (non-response, fast response, non-completion), using the pscl package, to determine whether test setting predicted hard failure of embedded performance validity metrics.

Construct Validity of the Points-Based EEfRT

Variables used to determine whether points impacted effort-based decision-making in this task included the outcome of hard task choice, coded as a zero when participants selected an easy task and one when participants selected a hard task. The reward predictor was not binned into high, medium, and low as some previous studies have done due to the expected continuous nature of the effect of this variable. Online test setting was coded as a zero for in-person and a one for online. All predictors, reward, probability, and online test setting were zero centered and scaled then hard task choice was regressed on the main effect of reward, the main effect of probability, the main effect of online vs. in-person test setting, and the interactions between reward and probability, reward and test setting, probability and test setting, and the three-way interaction of reward, probability, and test setting. After zero-centering and scaling the dependent variable, proportion of hard tasks chosen, as well as the predictors, reward, probability, and online test setting, multiple linear regression was conducted, using the lme4 package in R to determine whether significant main effects and interactions exist for reward, probability, and online test setting.

Profiles of Effort Expenditure

Latent profile analysis (LPA), a type of mixture modeling, is a method that enables heterogenous data to be separated into homogeneous groups (Robertson & Kaptein, 2016) therefore revealing the nomothetic span or patterns of relationships between the EEfRT and other cognitive tests. Latent profile analysis was used to explore the patterns of performance validity metrics across six tasks: EEfRT, Sternberg, Flanker, N-back, CPT, and PLT. First, non-response variables were created for each task. A second validity variable was also created for tasks where over-response might indicate low effort due to non-responses being correct answers on some trials (CPT and N-Back). All variables were z-scored before use in the latent profile analysis. Each model was checked visually and with a skewness statistic for normality.

The optimal number of profiles was determined in step-wise fashion, starting with a 1-profile model and continuing until model fit statistics failed to improve (Beauchaine, 2003). Current guidelines recommend using Akaike's information criteria (AIC), Bayesian information criteria (BIC), sample-size adjusted BIC (SABIC), with lower statistics indicating better fit, and entropy, with values closer to one as indicating better accuracy of profile classification, as well as checking for conceptual and qualitative reasonability in the profile patterns (Jason & Glenwick, 2016) to determine which model has the best fit (AIC, BIC, SABIC), best accuracy of profile classification (entropy) and interpretability (qualitative assessment). LPA was performed in R using the tidyLPA package (Rosenberg et al., 2019).

Three LPAs were fitted to the data, each following the stepwise procedure described above, in order to account for the possibility of multiple low-effort strategies (in the case of the CPT and N-Back) and to explore whether the proportion of hard tasks chosen on the EEfRT acts as an indicator of a low-effort profile. The first LPA, referred to as the "response" model, consisted of the z-scores for non-response on the CPT, EEfRT, Flanker, N-Back, PLT, and Sternberg. The second model, referred to as the "responsivity" model, included alternative scoring procedures for the CPT and N-Back. Since the CPT and N-Back both require non-responses as a correct answer, alternative scores were calculated (described below) to account for the possibility that over-responding could also be an indication of low effort. The final model, referred to as the "blend" model, aimed to explore whether the proportion of hard tasks chosen on the EEfRT relates to traditional effort scores, specifically non-response, on other cognitive tasks, the best fitting model from the previously described models was used, but the EEfRT score used was the z-score for percentage of hard tasks chosen instead of non-response.

For each task, non-response, or responsivity scores were calculated and checked for skewness both visually and with adjusted Fisher-Pearson coefficients of skewness and classified as normal (-0.5 to +0.5), moderately skewed (-1.0 to -0.5, +0.5 to +1.0) or high skewed (< - 1.0 or < + 1.0) (Bulmer, 1979; Piovesana & Senior, 2018).

Results

Participants

Of the 402 participants who consented to complete the study on Prolific.com, 393 completed the initial Qualtrics assessment that confirmed demographic characteristics and asked other questions about their cognitive and mental health. Of these, 342 completed all six tasks.

Of the 100 in-person participants who consented to complete the study, 55 were diagnosed with schizophrenia during the diagnostic interview and not included in the sample analyzed in this study. Of the 45 remaining participants, 19 participants had not completed the EEfRT at the time of writing this document leaving 26 in-person participants who completed all parts of the study and had data available for analysis. See Table 1 for participant demographics. While there were no significant differences between online and in-person participants in terms of age or sex, there was a significant difference in racial identities, with the online sample having far more non-White participants.

Embedded Performance Validity Tests

Table 2 reports multiple linear regression and zero-inflated Poisson model results for embedded performance validity measures. The count model for non-response during the choice phase had a significant positive effect for online setting ($b = 2.13$, $SE = 0.97$, $p = 0.03$) meaning that among participants who did have non-responses, online participants tended to have more non-responses than in-person. The fast response (responses less than 600ms) count model had a significant effect for online setting ($b = 0.92$, $SE = 0.46$, $p = 0.05$), meaning that among participants with at least one fast response (response to the decision phase faster than 600ms), online participants had more fast responses. The non-completion of the button pressing task count model had a significant positive effect for online setting ($b = 0.88$, $SE = 0.31$, $p < .005$), meaning that among participants with one or more incomplete button pressing tasks, online participants had more incomplete button pressing tasks than in-person. There was no significant effect of test setting on whether participants chose a mix of easy and hard tasks, nor was there a significant effect for test setting or trial number on hard-task choice. All participants pressed the spacebar at least once for every button pressing task.

Validity of the Points-Based EEfRT

Table 3 reports results for the multiple linear regression models predicting proportion of hard tasks chosen from reward (number of points), probability of receiving the reward, and test setting. Figure 2 visualizes the pattern of increasing proportion of hard tasks chosen with increasing reward and probability levels. The fixed effects revealed several significant predictors. There was a significant positive effect of reward (b = 0.59, SE = 0.03, 95% CI 0.5, 0.65, $p < .001$), meaning that as reward increased, proportion of hard tasks chosen also increased. There was also a significant positive effect of probability (b = 1.09, SE = 0.03, 95% CI 1.02, 1.15, $p < .001$) meaning that as probability increased, proportion of hard tasks chosen also increased. There were two significant interactions, a positive reward by probability interaction (b = 0.15, SE = 0.03, 95% CI 0.09, 0.21, $p < .001$) also known as expected value, meaning that the proportion of hard tasks chosen was highest when reward and probability level were both high, and a negative reward by online task setting interaction (b = -0.08, SE = 0.03, 95% CI -0.13, -0.03, $p < .005$), meaning that online participants chose even less hard tasks at the lowest reward levels compared to in-person participants.

Latent Profile Analyses

Table 4 reports fit statistics for the LPA models. The best fitting model (lowest AIC, BIC, and SABIC) among the response models was the 2-profile solution. The best fitting model among the responsivity models was the 3-profile solution for responsivity. The best fitting model among the blend models was the 2-profile solution. All the best fitting models also had strong support for accuracy of classification, with entropy values of 1.00.

Figure 3 shows the means for the 2-profile solution for the response model. Notably, one profile had means with high rates of response (n = 328) while the other profile had various levels of non-response on the cognitive tasks (n = 14), with the Flanker having the most extreme value with a z-score of -4.02. Figure 4 shows the means for the 3-profile solution for the responsivity model. This model included variables that account for over-responding as low-effort responses on the CPT and N-Back. The first profile has higher means of responsivity than both the second and third profiles. This could mean that the first profile classified participants with adequate response while the second and third profiles represent reduced effort on all tasks. In the second profile, the CPT has the lowest mean with all tasks other than the CPT having means in between those of the other two profiles. In the third profile, the CPT mean is between the CPT mean of the other two profiles, but means for all tasks other than the CPT are much lower than the other two profiles. In this solution, 14 participants were classified into the low-effort profile, 42 participants were classified into the abnormal CPT profile, and 287 participants were classified into the adequate effort profile. Figure 5 shows the means for the 2-profile solution for the blend of responsivity with the score for hard tasks chosen in the EEfRT. In this model one profile classifies 295 participants with adequate effort on all tasks while the other profile classifies 47 participants with low effort. In this case, the CPT has the most extreme low mean with a z-score value of -1.95. Importantly, this profile classified more than 5% of the sample into one of the profiles, making the results more qualitatively interpretable.

Discussion

The overarching goal of this study was to investigate the validity of a points-based version of the EEfRT and to contribute a possible solution to ethical and practical concerns with using a monetary version. The preliminary aim compared embedded performance validity tests from the EEfRT between online and in-person test setting to determine whether more invalid data was likely for this task in an online test setting. It was hypothesized that embedded performance validity tests would reveal more invalid data online, compared to the in-person test setting; specifically, online participants would demonstrate lower effort evidenced by increased odds of non-response rate, more fast-responses, more incomplete trials, and lower proportion of hard tasks chosen. The count model for non-response during the choice phase had a significant effect for online setting meaning that of the people who had non-responses, online participants tended to have more non-responses. This supports the hypothesis that online test setting was related to lower effort on the EEfRT. Also, the fast response count model had a significant effect for online setting suggesting an increased odds of fast responses for online participants with one or more fast responses. This could indicate low effort or could also suggest a different strategy pattern used by participants online, such as automatically choosing easy or hard trials for every trial without reading the stimuli on the decision-making screen. On the other hand, this could mean that the distractions in the home environment changed effort-based decision-making, or effort expenditure in online testing is lower due to lack of social pressure from a researcher present in the room.

While the non-response, fast response, and non-completion count models indicate the possibility of lower engagement with the task among online participants, none of the zero-inflated models had significant effects for online test setting, meaning estimated odds of observing excess zeros in the online test setting is not significant. This could suggest that although the participants who expend low effort on the task tend to expend even less effort than those in-person, a similar amount of people in both test settings expend an adequate amount of effort if adequate effort is considered to be having zero non-responses or fast responses. This supports the use of the test online in the future and suggests that non-response, fast-response, and non-completion may be useful metrics of whether participants were engaging with the task enough for their performance to be considered valid. Additionally, there was no significant effect of test setting on whether participants chose a mix of easy and hard tasks, nor was there a significant effect of test setting or trial number on hard-task choice. Additionally, since there was not an increase in choosing easy tasks on later trials, this task may be robust against fatigue both online and in-person, at least for the duration of 24 trials used with these participants.

The first aim of this study was to determine was to determine whether the reward and probability manipulations in the points-based version of the EEfRT impacted decision-making in-person and online in order to determine the construct validity of the points-based version. The pattern of response in which increasing reward and probability increase the proportion of hard tasks chosen is critical for understanding construct validity so that this pattern in healthy controls may be compared to abnormal patterns of reduced effort expenditure in participants with neurocognitive disorders. In the EEfRT, reward and probability are meant to influence effort-based decision making such that more hard tasks are chosen at higher reward levels so that effort valuation and expected value scores can be used to better understand the construct of effort. If this version of the task is truly measuring effort expenditure for reward, the reward and probability manipulations should work to increase proportion of hard tasks chosen at the highest reward and probability levels. To investigate whether the reward and probability manipulations were valid in the points-based version of the EEfRT, hard task choice was regressed onto reward and probability. It was hypothesized that participants would choose hard tasks more often at higher levels of reward and probability. Similar to monetary versions of the EEfRT, the points-based version showed significantly more hard tasks chosen at higher levels of reward and probability and a significant interaction between reward and probability in-person and online. The interaction suggests that the relationship between reward and hard task choice was less positive in-person compared to online.

The impact of reward on decision-making is of particular interest because this evidence of increased effort for points with no monetary value suggests that points may be a useful replacement for monetary reward in the EEfRT, making the task more accessible to researchers without large funds or who have ethical concerns about paying subjects based on performance. Increased effort for points also supports the potential avenue of gamification of cognitive tasks to incentivize performance on cognitive tests that are not typically associated with high-stakes outcomes, such as those used in research with undergraduate populations. Further research must be conducted to determine whether populations with deficits in reward valuation also exhibit similar patterns of behavior when rewarded with points. The impact of probability on decision-making also supports the use of the points-based version for investigating effort-based decision-making because increasing effort at higher levels of probability is a key component of reward valuation. If this task had not shown a significant increase in hard tasks chosen at higher probability levels, the manipulation could not be informative of reward valuation. The significant interaction of reward and probability suggests that expected value can also be investigated with the points-based version of the EEfRT. Since the three most important manipulations of effort-based decision-making had significant effects for increasing effort expenditure in the points-based version, this suggests that this task could be used alongside other cognitive tests to better understand possible components of effort that could underlie low test scores without the complications of paying for performance. Although more research must be done to determine whether the EEfRT could be useful in a clinical context as a standalone validity tests, the results of this study support the use of the points-based EEfRT for investigating the subcomponents of effort-based decision-making, at least in a healthy sample. The hypothesis that reward, probability, and expected value would increase hard-task choices was supported.

There was also a significant interaction between test setting and reward suggesting that online participants were less likely to work for lower point levels than in-person participants. This does not support the hypothesis that the relationship between reward and hard-task choice would be attenuated online. In fact, the relationship found in these data was in the opposite direction with points having more impact in the online test setting. While this could suggest that participants online are more sensitive to low levels of reward, this finding could also be an artifact of the small sample size in-person. Further research is needed to determine whether points impact decision-making differently in different test environments.

The second aim of this study was to determine whether latent groupings of participants based on performance validity metrics exist across different cognitive tasks. Three latent profile models using response scores, responsivity scores (that accounted for the possibility of over-response being a low-effort strategy), and a blend of responsivity scores and the proportion of hard tasks chosen on the EEfRT revealed profiles of adequate effort and low effort in all three cases. Utilizing three latent profile analyses with differing variables enabled exploration of the pattern of significant relationships, or nomothetic span, across measures of different cognitive domains.

The best fitting response latent profile analysis was a 2-profile solution and classified 14 participants into a low-effort profile and the rest of the participants into an adequate-effort profile. Due to less than 5% of the sample being classified into one of the profiles, these results should be interpreted with caution. One interpretation could be that most of the participants in the sample engaged with the task at an adequate level with relatively few participants non-responding at a high rate. The EEfRT response mean was over one standard deviation below the average, suggesting that non-response on the EEfRT could be an indicator of low effort across other cognitive tasks.

The best fitting responsivity latent profile analysis was a 3-profile solution and classified 14 participants into a low-effort profile, 42 participants into an abnormal CPT profile, and the rest of the participants into an adequate effort profile. As in the response model, the low-effort profile had an EEfRT mean over one standard deviation below the group average. The means for the abnormal CPT profile were close to the average for all tasks except for the CPT.

The best fitting blend latent profile analysis was a 2-profile solution and classified 47 participants into a low-effort profile and the rest into an adequate-effort profile. When the proportion of hard tasks score was used instead of response on the EEfRT, the pattern of the low-effort profile changed. In this case, the means for most tasks are within one standard deviation of the adequate-effort profile means and the CPT has the lowest mean for responsivity.

The results of this study should be interpreted in light of several limitations. First, small in-person sample size implies that analyses were not powered to detect smaller effects. A second limitation was that online participants were paid less per hour than in-person participants. More investigation is needed to determine whether participants paid more per hour in an online setting expend more effort on cognitive tests. Although the points-based version was not directly compared to a monetary version of this task, the points-based reward did show patterns of effect on effort expenditure similar to the patterns in previous monetary versions. A direct comparison between a points-based version and monetary version would provide more support for convergent validity. Another useful direct comparison could be associating EEfRT scores with performance on stand-alone performance validity metrics like the test of memory malingering (Tombaugh, 1997). Supporting the convergent validity of this measure by direct comparison to a monetary version and standalone performance validity tests could further validate the dual nature of this task as a potential effort-based decision-making task and performance validity task. On the other hand, if a strong relationship is not found between the EEfRT and either decision-making or performance validity tests, that might suggest divergent validity of the EEfRT as a measure of a unique effort-related construct.

The physical nature of the task is a limitation for multiple reasons. First, no calibration was conducted to ensure that motor-performance did not impact willingness to expend effort. Second, due to the lack of supervision, we cannot be sure that online participants truly used their dominant index finger for easy tasks and non-dominant pinky finger for hard tasks. Third, previous research has indicated both shared and distinct neural mechanisms for physical and cognitive effort which could suggest that a physical effort task might not be the best indicator of cognitive effort-based decision-making (L. Schmidt et al., 2012)

Overall, results support the use of points as a valid reward for manipulating effort-based decision-making within the context of measuring effort as a construct. Online and in-person participants provided generally valid data with relatively few participants having extreme patterns of non-response or low-effort responses (indicated by fast reaction times and non-completion). Additionally, reward and probability manipulations worked to increase hard-task choices online and in-person suggesting that effort-based decision-making can be investigated without the ethical and practical concerns of paying for performance on the EEfRT. While the EEfRT has been suggested to be an “objective” measure of effort (Treadway et al., 2009), it is not yet clear if clinicians could include this test in a cognitive battery to determine performance validity on other cognitive tests.

One of the challenges of this study is the potential differences between measuring effort as a construct and adequate effort expenditure in the context of performance validity. Although both concepts are referred to in terms of “effort expenditure”, it is not yet well-understood if effort-based decision-making and effort required to engage in a cognitive task relate to the same psychological construct and neuropsychological mechanisms. While effort as a construct is often considered in terms of group differences (e.g., reduced effort expenditure in participants with schizophrenia), individual differences may also reveal important aspects of reward processing.

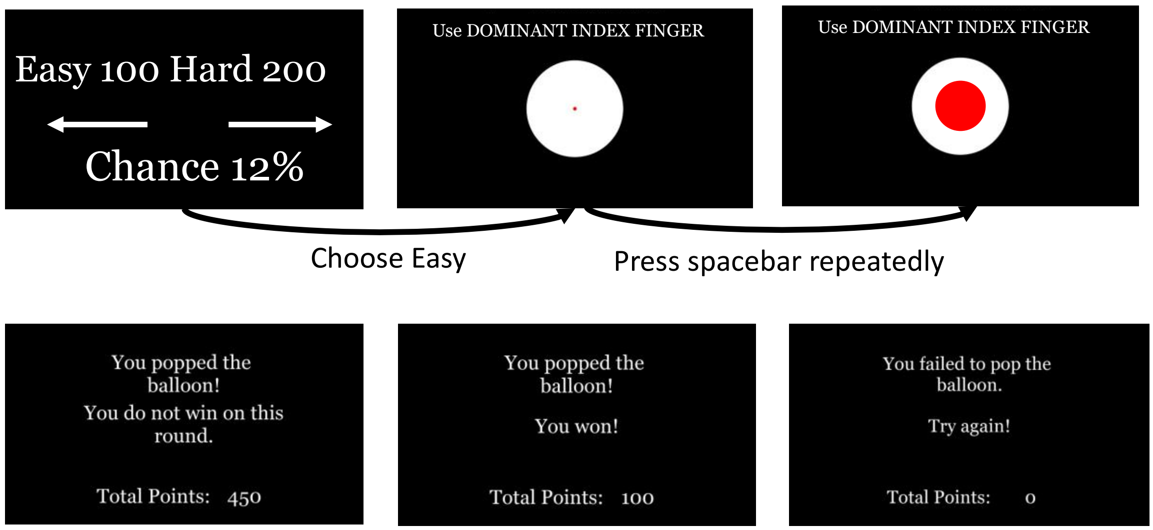

Figure 1.

Points-Based Effort Expenditure for Rewards Task (EEfRT)

Note. The Effort Expenditure for Rewards Task (EEfRT) offered participants a choice between an easy task and a hard task. After participants made a choice, they were required to press a button (easy = 30 times in 7 seconds, hard = 100 times in 20 seconds) to increase the size of the red circle until the red circle became as large as the white circle, or "popping the balloon". Participants received one of three feedback options depending on their performance and the probability of receiving the reward.

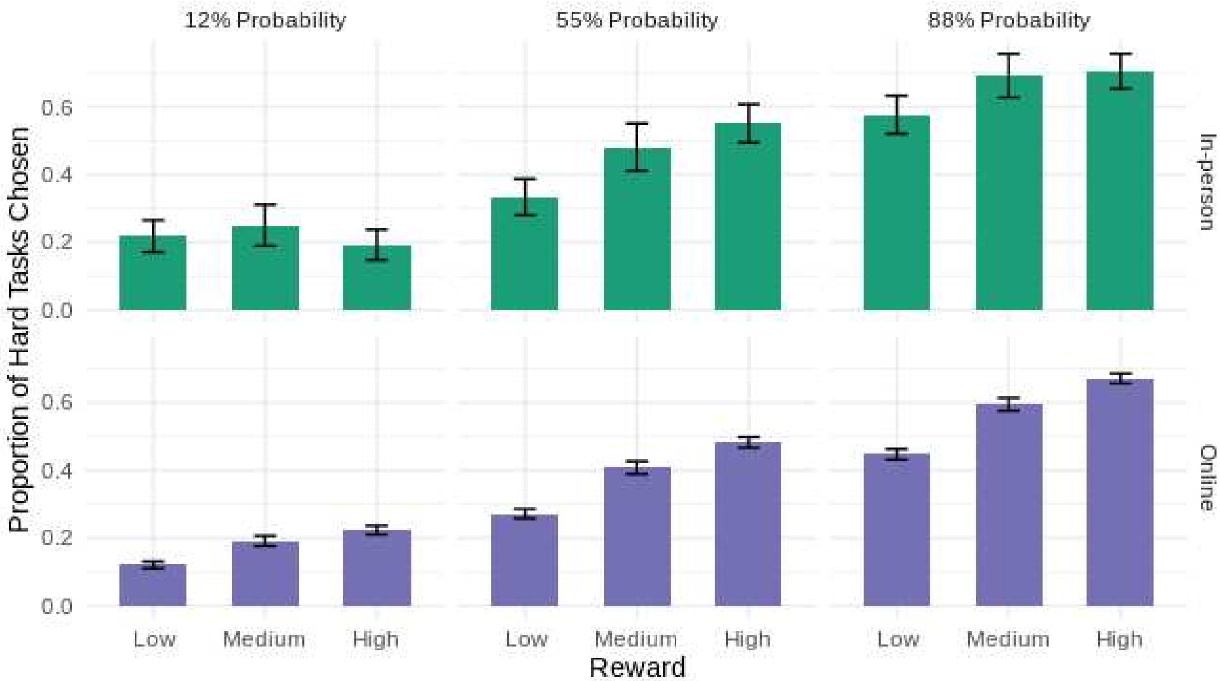

Figure 2.

Proportion of Hard Tasks Chosen by Probability and Reward

Note. Bar chart of proportion of hard tasks chosen with the top of the bar representing the average proportion of hard tasks chosen calculated by probability (12%, 55%, and 88%), and reward (low = 150, 250, 350 points; medium = 450, 550 points; high = 650, 750, 800 points) for in-person (green) and online (blue) participants, with standard error indicated by black error bars.

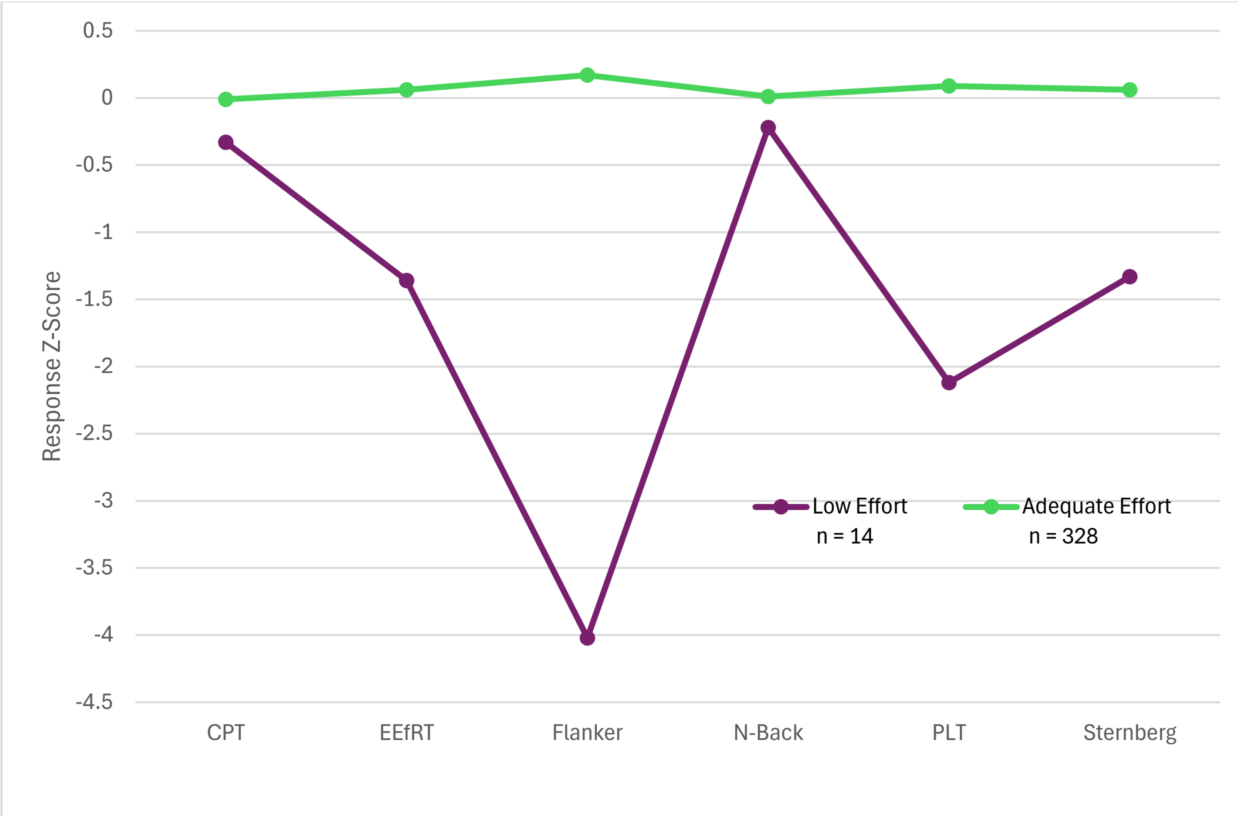

Figure 3.

2-Profile Solution Means for Response

Note. Latent profile analysis plot of mean response rates for all tasks. Notably, one profile had means with high rates of response (green) while the other profile had various levels of non-response on the cognitive tasks (purple). Continuous Pointing Task (CPT), Effort-Expenditure for Rewards Task (EEfRT), Probabilistic Learning Task (PLT).

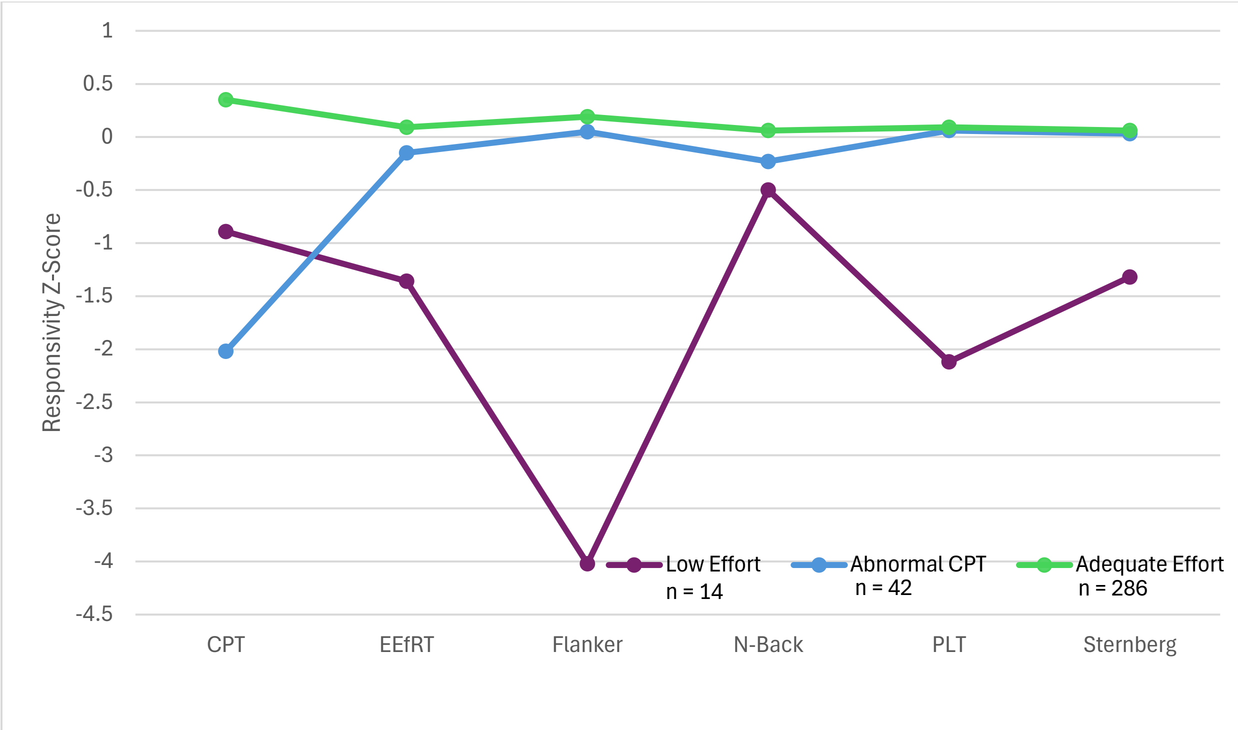

Figure 4.

3-Profile Solution Means for Responsivity

Note. Latent profile analysis plot of mean response rates for the EEfRT, Flanker, PLT, and Sternberg, similar to Figure 3. For the CPT and N-Back, response rate was replaced with a responsivity score calculated so that over-responding and under-responding are both represented by a lower score due to the nature of the task involving non-response as a correct answer. Notably, one profile had means with high rates of response (green), another profile had various levels of non-response on the cognitive tasks (purple), and a final profile had high rates of response on most tasks with a low responsivity mean for the CPT (blue). Continuous Pointing Task (CPT), Effort-Expenditure for Rewards Task (EEfRT), Probabilistic Learning Task (PLT).

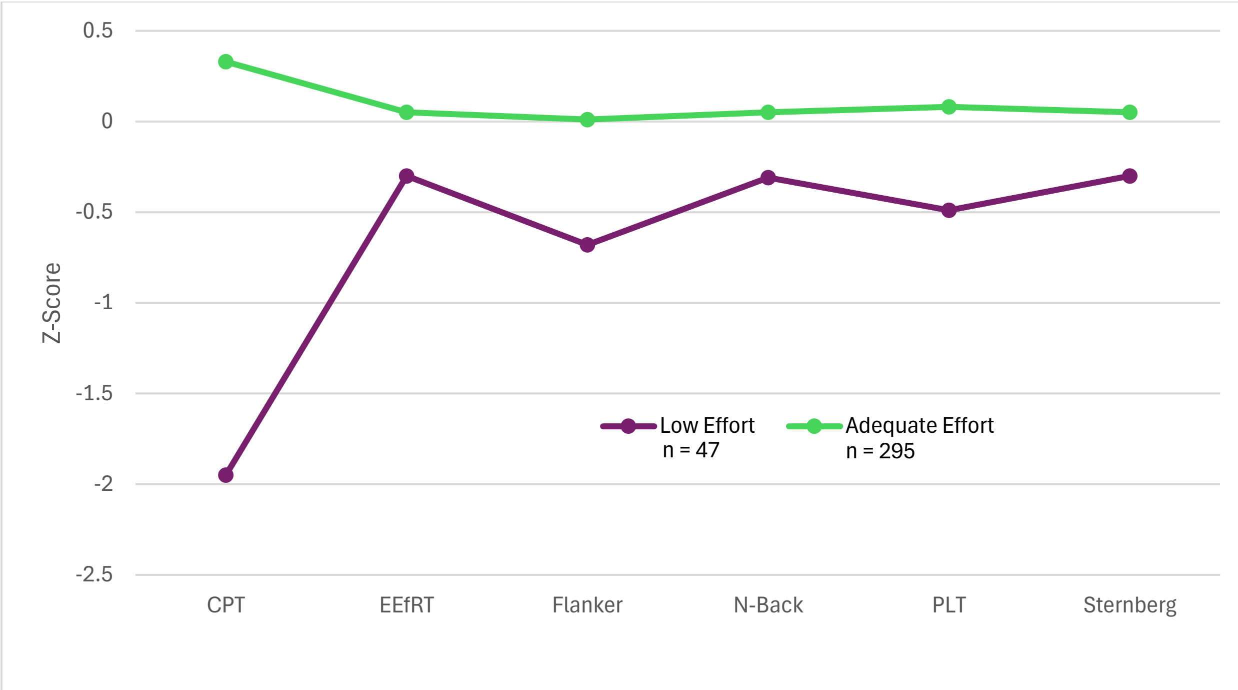

Figure 5.

2-Profile Solution Means for Blend of EEfRT Hard-Task Choices with Non-Response

Note. Latent profile analysis plot of mean response rates for the Flanker, PLT, and Sternberg, similar to the Figures 3 and 4. Similar to Figure 4, responsivity scores were used for CPT and N-Back, such that over- and under- responding are reflected by a lower score. The score for the EEfRT is proportion of hard tasks chosen. Notably, one profile had means with high rates of response (green) while the other profile had various levels of non-response on the cognitive tasks (purple). Continuous Pointing Task (CPT), Effort-Expenditure for Rewards Task (EEfRT), Probabilistic Learning Task (PLT).

Table 1

Participant Demographics

	In-Person (N = 26)	Online (N = 342)	p value
Age (years)			
Mean (SD)	42.3 (14.6)	39.9 (14.5)	0.426
Median [Min, Max]	38.0 [20.0, 68.0]	37.0 [19.0, 74.0]	
Sex			
Male	14 (53.8%)	166 (48.5%)	0.75
Female	12 (46.2%	176 (51.5%)	
Race			< 0.001
Asian	0 (0%)	54 (15.8%)	
American Indian/Alaska Native	0 (0%)	10 (2.9%)	
Black or African American	1 (3.8%)	60 (17.5%)	
More than one race	3 (11.5%)	37 (10.8%)	
Unknown	1 (3.8%)	0 (0%)	
White	21 (80%)	181 (52.9%)	

Table 2.
Note. Number of participants = 368. CI = confidence interval; LL = lower limit; UL = upper limit.
a 0 = in-person, 1 = online; b = trial number out of 23 total trials.

EEfRT Embedded Performance Validity Tests

Embedded Performance Validity Test	Effect		Estimate	SE	95% CI		*p*
					LL	*UL*	
Non-response							
	Count model						
		Intercept	-1.15	0.97	-3.06	0.76	0.24
		Online[a]	2.13	0.97	0.22	4.04	0.03
	Zero-inflation model						
		Intercept	-1.74	6.01	-13.51	10.04	0.77
		Online[a]	2.50	6.01	-9.28	14.28	0.68
Fast response							
	Count model						
		Intercept	0.47	0.46	-0.43	1.37	0.31
		Online[a]	0.92	0.46	0.01	1.82	0.05
	Zero-inflation model						
		Intercept	1.43	0.62	0.22	2.64	0.02
		Online[a]	-0.38	0.63	-1.61	0.85	0.54
Non-completion							
	Count model						
		Intercept	0.47	0.31	-0.13	1.07	0.13
		Online[a]	0.88	0.31	0.28	1.49	< 0.005
	Zero-inflation model						
		Intercept	0.26	0.52	-0.76	1.29	0.62
		Online[a]	-0.44	0.54	-1.49	0.61	0.41
Mix of choices							
	Fixed effects						
		Intercept	-0.32	0.28	-0.85	0.21	0.25
		Online[a]	-0.37	0.29	-0.95	0.19	0.20
	Random effects						
		Within-participant variance	1.779	1.334			
Fatigue							
	Fixed effects						
		Intercept	-0.14	0.31	-0.83	0.49	0.65
		Online[a]	-0.25	0.32	-0.88	0.50	0.21
		Trial[b]	-0.25	0.33	-0.04	0.01	0.45
		Online[a] * trial interaction	-0.01	0.01	-0.04	0.02	0.41
	Random effects						
		Within-participant variance	1.81	1.34			

Table 3.

Note. Number of observations = 8832, number of participants = 368. CI = confidence interval; *LL* = lower limit; *UL* = upper limit.

[a] 150 – 850 points in increments of 100. [b] = 12%, 50%, 88%. [c] 0 = in-person, 1 = online.

Hard Task Choice Predicted by Reward, Probability, and Test Setting

Effect	Estimate	*SE*	95% CI		*p*
			LL	*UL*	
Fixed effects					
Intercept	-0.85	0.10	-1.04	-0.65	< 0.001
Reward[a]	0.59	0.03	0.53	0.65	< 0.001
Probability[b]	1.09	0.03	1.02	1.15	< 0.001
Online[c]	0.13	0.10	-0.06	0.32	0.18
Reward*probability interaction	0145	0.03	0.09	0.21	< 0.001
Reward*online interaction	-0.08	0.03	-0.13	-0.03	< 0.005
Probability*online interaction	-0.03	0.03	-0.09	0.02	0.27
Random effects					
Within participant variance	3.09	1.75			

Table 4.
Note. AIC = Akaike Information Criterion; BIC = Basian Information Criterion, SABIC = Sample size-adjusted Bayesian Information Criterion, * lowest value among models.

Model Fit Indices for the Latent Profile Solutions

Model	AIC	BIC	SBAIC	Entropy
Response, 1- profile	5841	5887	5849	1.00
Response, 2 - profiles	5451*	5523*	5463*	1.00
Response, 3 - profiles	5465	5565	5483	0.44
Responsivity, 1- profile	5841	5887	5849	1.00
Responsivity, 2 - profiles	5366	5439	5379	1.00
Responsivity, 3 - profiles	5318*	5417*	5335*	0.95
Responsivity, 4 - profiles	5332	5458	5354	0.63
Blend with responsivity, 1 – profile	5840	5887	5848	1.00
Blend with responsivity, 2 – profiles	5682*	5755*	5695*	0.90
Blend with responsivity, 3 - profiles	5691	5791	5709	0.60

REFERENCES

Abeare, K., Cutler, L., An, K. Y., Razvi, P., Holcomb, M., & Erdodi, L. A. (2022). BNT–15: Revised performance validity cutoffs and proposed clinical classification ranges. *Cognitive and Behavioral Neurology*, *35*(3), 155. https://doi.org/10.1097/WNN.0000000000000304

Assadi, S. M., Yücel, M., & Pantelis, C. (2009). Dopamine modulates neural networks involved in effort-based decision-making. *Neuroscience & Biobehavioral Reviews*, *33*(3), 383–393. https://doi.org/10.1016/j.neubiorev.2008.10.010

Bailey, K. C., Soble, J. R., & O'Rourke, J. J. F. (2018). Clinical utility of the Rey 15-Item Test, recognition trial, and error scores for detecting noncredible neuropsychological performance in a mixed clinical sample of veterans. *The Clinical Neuropsychologist*, *32*(1), 119–131. https://doi.org/10.1080/13854046.2017.1333151

Banerjee, M. (2020). An exploratory study of online equity: Differential levels of technologica access and technological efficacy among underserved and underrepresented student populations in higher education. *Inderdisciplinary Journal of E-Skills and Lifelong Learning*, *16*, 93–121.

Barch, D. M., Treadway, M. T., & Schoen, N. (2014). Effort, anhedonia, and function in schizophrenia: Reduced effort allocation predicts amotivation and functional impairment. *Journal of Abnormal Psychology*, *123*(2), 387–397. psyh. https://doi.org/10.1037/a0036299

Beauchaine, T. P. (2003). Taxometrics and developmental psychopathology. *Development and Psychopathology*, *15*(3), 501–527. https://doi.org/10.1017/S0954579403000270

Behavioral Assessment Methods for RDoC Constructs—National Institute of Mental Health (NIMH). (2023). https://www.nimh.nih.gov/about/advisory-boards-and-groups/namhc/reports/behavioral-assessment-methods-for-rdoc-constructs

Bekk, M., Eppmann, R., Klein, K., & Völckner, F. (2022). All that glitters is not gold: An investigation into the undesired effects of gamification and how to mitigate them through gamification design.

International Journal of Research in Marketing, *39*(4), 1059–1081. https://doi.org/10.1016/j.ijresmar.2022.03.002

Belleville, S., LaPlume, A. A., & Purkart, R. (2023). Web-based cognitive assessment in older adults: Where do we stand? *Current Opinion in Neurology*, *36*(5), 491–497. https://doi.org/10.1097/WCO.0000000000001192

Berntzen, L., Böhm, S., & International Academy, Research, and Industry Association (Eds.). (2013). *CENTRIC 2013: The Sixth International Conference on Advances in Human Oriented and Personalized Mechanisms, Technologies, and Services: October 27-November 1, 2013, Venice, Italy*. Curran Associates, Inc.

Bigler, E. D. (2012). Symptom validity testing, effort, and neuropsychological assessment. *Journal of the International Neuropsychological Society*, *18*(4), 632–640. https://doi.org/10.1017/S1355617712000252

Bigler, E. D. (2014). Effort, symptom validity testing, performance validity testing and traumatic brain injury. *Brain Injury*, *28*(13–14), 1623–1638. https://doi.org/10.3109/02699052.2014.947627

Bijleveld, E., van Breukelen, F. N., de Segovia Vicente, D., & Schutter, D. J. L. G. (2023). Mapping the dose–response relationship between monetary reward and cognitive performance. *Motivation Science*, *9*(3), 205–215. https://doi.org/10.1037/mot0000299.supp

Bismark, A. W., Thomas, M. L., Tarasenko, M., Shiluk, A. L., Rackelmann, S. Y., Young, J. W., & Light, G. A. (2018). Relationship between effortful motivation and neurocognition in schizophrenia. *Schizophrenia Research*, *193*, 69–76. https://doi.org/10.1016/j.schres.2017.06.042

Boynton, M., Portnoy, D. B., & Johnson, B. T. (2013). Exploring the ethics and psychological impact of deception in psychological research. *IRB*, *35*(2), 7–13.

Brown, J. K., Waltz, J. A., Strauss, G. P., McMahon, R. P., Frank, M. J., & Gold, J. M. (2013). Hypothetical decision making in schizophrenia: The role of expected value computation and “irrational” biases. *Psychiatry Research*, *209*(2), 142–149. https://doi.org/10.1016/j.psychres.2013.02.034

Bruya, B., & Tang, Y.-Y. (2018). Is attention really effort? Revisiting Daniel Kahneman's influential 1973 book attention and effort. *Frontiers in Psychology*, *9*, 1133. https://doi.org/10.3389/fpsyg.2018.01133

Bulmer, M. (1979). *Principles of statistics*. Dover Publications.

Byrne, K. A., Lockwood, P. L., Ghaiumy Anaraky, R., & Liu, Y. (2023). Age differences in prosocial behavior depend on effort costs. *The Journals of Gerontology: Series B*, *78*(6), 948–958. https://doi.org/10.1093/geronb/gbac194

Chafetz, M. D., Prentkowski, E., & Rao, A. (2011). To work or not to work: Motivation (not low IQ) determines symptom validity test findings. *Archives of Clinical Neuropsychology*, *26*(4), 306–313. https://doi.org/10.1093/arclin/acr030

Chafetz, M. D., Williams, M. A., Ben-Porath, Y. S., Bianchini, K. J., Boone, K. B., Kirkwood, M. W., Larrabee, G. J., & Ord, J. S. (2015). Official position of the American Academy of Clinical Neuropsychology Social Security Administration policy on validity testing: Guidance and recommendations for change. *The Clinical Neuropsychologist*, *29*(6), 723–740. https://doi.org/10.1080/13854046.2015.1099738

Chaldogeridis, A., & Tsiatsos, T. (2022). Gamification Techniques and Best Practices in Computerized Working Memory Training: A Systematic Literature Review. *Applied Sciences*, *12*(19), Article 19. https://doi.org/10.3390/app12199785

Chong, T. T.-J., Apps, M., Giehl, K., Sillence, A., Grima, L. L., & Husain, M. (2017). Neurocomputational mechanisms underlying subjective valuation of effort costs. *PLOS Biology*, *15*(2), e1002598. https://doi.org/10.1371/journal.pbio.1002598

Culbreth, A. J., Dershwitz, S. D., Barch, D. M., & Moran, E. K. (2023). Associations Between Cognitive and Physical Effort–Based Decision Making in People With Schizophrenia and Healthy Control Subjects. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(7), 695–702. https://doi.org/10.1016/j.bpsc.2023.02.003

Culbreth, A. J., Moran, E. K., & Barch, D. M. (2018a). Effort-based decision-making in schizophrenia. *Current Opinion in Behavioral Sciences*, *22*, 1–6. https://doi.org/10.1016/j.cobeha.2017.12.003

Culbreth, A. J., Moran, E. K., & Barch, D. M. (2018b). Effort-cost decision-making in psychosis and depression: Could a similar behavioral deficit arise from disparate psychological and neural mechanisms? *Psychological Medicine*, *48*(6), 889–904. https://doi.org/10.1017/S0033291717002525

Culbreth, A. J., Westbrook, A., Braver, T. S., & Barch, D. M. (2020). Effort in daily life: Relationships between experimental tasks and daily experience. *Motivation Science*, *6*(3), 303–308. psyh. https://doi.org/10.1037/mot0000159

de Winter, J. C. F. (2014). Controversy in human factors constructs and the explosive use of the NASA-TLX: A measurement perspective. *Cognition, Technology & Work*, *16*(3), 289–297. https://doi.org/10.1007/s10111-014-0275-1

Denny, P. (2013). The effect of virtual achievements on student engagement. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 763–772. https://doi.org/10.1145/2470654.2470763

Emanuel, E., Abdoler, E., & Stunkel, L. (2016). Research ethics: How to treat people who participate in research. In A. E. Kazdin (Ed.), *Methodological issues and strategies in clinical research (4th ed.).* (pp. 513–523). American Psychological Association. https://doi.org/10.1037/14805-031

Engström, M., Landtblom, A.-M., & Karlsson, T. (2013). Brain and effort: Brain activation and effort-related working memory in healthy participants and patients with working memory deficits. *Frontiers in Human Neuroscience*, *7*. https://doi.org/10.3389/fnhum.2013.00140

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Feenstra, H. E. M., Murre, J. M. J., Vermeulen, I. E., Kieffer, J. M., & Schagen, S. B. (2018). Reliability and validity of a self-administered tool for online neuropsychological testing: The Amsterdam

Cognition Scan. *Journal of Clinical and Experimental Neuropsychology*, *40*(3), 253–273. https://doi.org/10.1080/13803395.2017.1339017

Fleming, H., Robinson, O. J., & Roiser, J. P. (2023). Measuring cognitive effort without difficulty. *Cognitive, Affective, & Behavioral Neuroscience*, *23*(2), 290–305. https://doi.org/10.3758/s13415-023-01065-9

Geaney, J. T., Treadway, M. T., & Smillie, L. D. (2015). Trait Anticipatory Pleasure Predicts Effort Expenditure for Reward. *PLOS ONE*, *10*(6), e0131357. https://doi.org/10.1371/journal.pone.0131357

Green, P. (2007). The Pervasive Influence of Effort on Neuropsychological Tests. *Physical Medicine and Rehabilitation Clinics of North America*, *18*(1), 43–68. https://doi.org/10.1016/j.pmr.2006.11.002

Green, P., Rohling, M. L., Lees-Haley, P. R., & Iii, L. M. A. (2001). Effort has a greater effect on test scores than severe brain injury in compensation claimants. *Brain Injury*, *15*(12), 1045–1060. https://doi.org/10.1080/02699050110088254

Greher, M. R., & Wodushek, T. R. (2017). Performance Validity Testing in Neuropsychology: Scientific Basis and Clinical Application—A Brief Review. *Journal of Psychiatric Practice*, *23*(2), 134–140. https://doi.org/10.1097/PRA.0000000000000218

Halperin, I., & Vigotsky, A. (2024). *A Conceptual Framework of Effort and Perception of Effort*. https://doi.org/10.31234/osf.io/h4dfe

Herrera, P. M., Speranza, M., Hampshire, A., & Bekinschtein, T. A. (2014). Monetary rewards modulate inhibitory control. *Frontiers in Human Neuroscience*, *8*, 257. https://doi.org/10.3389/fnhum.2014.00257

Hishamuddin Abdul Rahman, M., Ismail, D., Saria Binti Mohd Noor, A., & Syazwani Binti Mat Salleh, N. (2018). Gamification elements and their impacts on teaching and learning – A Review. *The International Journal of Multimedia & Its Applications*, *10*(6), 1–10.

Horan, W. P., Johnson, M. W., & Green, M. F. (2017). Altered experiential, but not hypothetical, delay discounting in schizophrenia. *Journal of Abnormal Psychology*, *126*(3), 301–311. https://doi.org/10.1037/abn0000249

Horne, S. J., Topp, T. E., & Quigley, L. (2021). Depression and the willingness to expend cognitive and physical effort for rewards: A systematic review. *Clinical Psychology Review*, *88*, 102065. https://doi.org/10.1016/j.cpr.2021.102065

Howieson, D. (2019). Current limitations of neuropsychological tests and assessment procedures. *The Clinical Neuropsychologist*, *33*(2), 200–208. https://doi.org/10.1080/13854046.2018.1552762

Hughes, D. M., Yates, M. J., Morton, E. E., & Smillie, L. D. (2015). Asymmetric frontal cortical activity predicts effort expenditure for reward. *Social Cognitive and Affective Neuroscience*, *10*(7), 1015–1019. https://doi.org/10.1093/scan/nsu149

Hutchinson, J. C., Sherman, T., Martinovic, N., & Tenenbaum, G. (2008). The Effect of Manipulated Self-Efficacy on Perceived and Sustained Effort. *Journal of Applied Sport Psychology*, *20*(4), 457–472. https://doi.org/10.1080/10413200802351151

Jason, L., & Glenwick, D. (2016). *Handbook of Methodological Approaches to Community-based Research: Qualitative, Quantitative, and Mixed Methods*. Oxford University Press.

Jin, X. (2023). The role of effort in understanding academic achievements: Empirical evidence from China. *European Journal of Psychology of Education*. https://doi.org/10.1007/s10212-023-00694-5

Keel, P. K., Kennedy, G. A., Rogers, M. L., Joyner, K. J., Bodell, L. P., Forney, K. J., & Duffy, M. E. (2022). Reliability and validity of a transdiagnostic measure of reward valuation effort. *Psychological Assessment*, *34*(5), 419–430. https://doi.org/10.1037/pas0001107

Kirchner, W. K. (1958). Age differences in short-term retention of rapidly changing information. *Journal of Experimental Psychology*, *55*(4), 352–358. https://doi.org/10.1037/h0043688

Kulas, J. F., Axelrod, B. N., & Rinaldi, A. R. (2014). Cross-Validation of Supplemental Test of Memory Malingering Scores as Performance Validity Measures. *Psychological Injury and Law*, *7*(3), 236–244. https://doi.org/10.1007/s12207-014-9200-4

Kurzban, R., Duckworth, A., Kable, J. W., & Myers, J. (2013). An opportunity cost model of subjective effort and task performance. *Behavioral and Brain Sciences*, *36*(6), 661–679. https://doi.org/10.1017/S0140525X12003196

Lance, C. E., Lance, C. E., & Vandenberg, R. J. (2010). *Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences*. Routledge.

Lauriola, M., Panno, A., Levin, I. P., & Lejuez, C. W. (2014). Individual Differences in Risky Decision Making: A Meta-analysis of Sensation Seeking and Impulsivity with the Balloon Analogue Risk Task. *Journal of Behavioral Decision Making*, *27*(1), 20–36. https://doi.org/10.1002/bdm.1784

Lejuez, C. W., Read, J. P., Kahler, C. W., Richards, J. B., Ramsey, S. E., Stuart, G. L., Strong, D. R., & Brown, R. A. (2002). Evaluation of a behavioral measure of risk taking: The Balloon Analogue Risk Task (BART). *Journal of Experimental Psychology: Applied*, *8*(2), 75–84. https://doi.org/10.1037/1076-898X.8.2.75

Leotti, L. A., & Wager, T. D. (2010). Motivational influences on response inhibition measures. *Journal of Experimental Psychology. Human Perception and Performance*, *36*(2), 430–447. https://doi.org/10.1037/a0016802

Lewis, Z. H., Swartz, M. C., & Lyons, E. J. (2016). What's the Point?: A Review of Reward Systems Implemented in Gamification Interventions. *Games for Health Journal*, *5*(2), 93–99. https://doi.org/10.1089/g4h.2015.0078

Lippa, S. M. (2018). Performance validity testing in neuropsychology: A clinical guide, critical review, and update on a rapidly evolving literature. *The Clinical Neuropsychologist*, *32*(3), 391–421. https://doi.org/10.1080/13854046.2017.1406146

Locey, M. L., Jones, B. A., & Rachlin, H. (2011). Real and hypothetical rewards. *Judgment and Decision Making*, *6*(6), 552–564.

Loeys, T., Moerkerke, B., De Smet, O., & Buysse, A. (2012). The analysis of zero-inflated count data: Beyond zero-inflated Poisson regression. *British Journal of Mathematical and Statistical Psychology*, *65*(1), 163–180. https://doi.org/10.1111/j.2044-8317.2011.02031.x

Lopez-Gamundi, P., & Wardle, M. C. (2018). The cognitive effort expenditure for rewards task (C-EEfRT): A novel measure of willingness to expend cognitive effort. *Psychological Assessment*, *30*(9), 1237–1248. https://doi.org/10.1037/pas0000563

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D., & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A Systematic review of applications and efficacy. *JMIR Serious Games*, *4*(2), e11. https://doi.org/10.2196/games.5888

Mækelæ, M. J., Klevjer, K., Westbrook, A., Eby, N. S., Eriksen, R., & Pfuhl, G. (2023). Is it cognitive effort you measure? Comparing three task paradigms to the Need for Cognition scale. *PLOS ONE*, *18*(8), e0290177. https://doi.org/10.1371/journal.pone.0290177

McGuire, C., Crawford, S., & Evans, J. J. (2019). Effort testing in dementia assessment: A Systematic review. *Archives of Clinical Neuropsychology*, *34*(1), 114–131. https://doi.org/10.1093/arclin/acy012

McWhirter, L., Ritchie, C. W., Stone, J., & Carson, A. (2020). Performance validity test failure in clinical populations—A systematic review. *Journal of Neurology, Neurosurgery & Psychiatry*. https://doi.org/10.1136/jnnp-2020-323776

Merritt, V. C., Rabinowitz, A. R., Guty, E., Meyer, J. E., Greenberg, L. S., & Arnett, P. A. (2019). Financial incentives influence ImPACT validity indices but not cognitive composite scores. *Journal of Clinical and Experimental Neuropsychology*, *41*(3), 312–319. https://doi.org/10.1080/13803395.2018.1551519

Millis, S. R. (2008). What clinicians really need to know about symptom exaggeration, insufficient effort, and malingering: Statistical and measurement matters. In *Neuropsychology of Malingering Casebook*. Psychology Press.

Miranda, A. T., & Palmer, E. M. (2014). Intrinsic motivation and attentional capture from gamelike features in a visual search task. *Behavior Research Methods*, *46*(1), 159–172. https://doi.org/10.3758/s13428-013-0357-7

Molinaro, G., & Collins, A. G. E. (2023). Intrinsic rewards explain context-sensitive valuation in reinforcement learning. *PLOS Biology*, *21*(7), e3002201. https://doi.org/10.1371/journal.pbio.3002201

Moore, P. S., Brifcani, S., & Worthington, A. (Eds.). (2021). *Neuropsychological aspects of brain injury litigation: A medicolegal handbook for lawyers and clinicians*. Routledge.

Morrissey, S., Gillings, R., & Hornberger, M. (2023). *Feasibility and reliability of online vs in-person cognitive testing in healthy older people* (p. 2023.07.05.23292229). medRxiv. https://doi.org/10.1101/2023.07.05.23292229

Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences*, *107*(49), 20911–20916. https://doi.org/10.1073/pnas.1013305107

National Advisory Mental Health Council Workgroup on Tasks and Measures for Research Domain Criteria (RDoC). (2016). *Behavioral assessment methods for rdoc constructs*. https://www.nimh.nih.gov/sites/default/files/documents/about/advisory-boards-and-groups/namhc/reports/rdoc_council_workgroup_report.pdf

National Telecommunications and Information Administration. (2023). *Digital Equity Act Programs*. BroadbandUSA. https://broadbandusa.ntia.doc.gov/funding-programs/digital-equity-act-programs

Nguyen, N. H., Albert, A. B., Orman, S. V., Forken, P., Blatt, S. D., Fremont, W. P., Faraone, S. V., & Glatt, S. J. (2019). Effort valuation and psychopathology in children and adults. *Psychological Medicine*, *49*(16), 2801–2807. https://doi.org/10.1017/S0033291718003884

Ohmann, H. A., Kuper, N., & Wacker, J. (2018). Left frontal anodal tDCS increases approach motivation depending on reward attributes. *Neuropsychologia*, *119*, 417–423. https://doi.org/10.1016/j.neuropsychologia.2018.09.002

Ohmann, H. A., Kuper, N., & Wacker, J. (2022). Examining the reliability and validity of two versions of the Effort-Expenditure for Rewards Task (EEfRT). *PLOS ONE*, *17*(1), e0262902. https://doi.org/10.1371/journal.pone.0262902

Okuhata, S., Kusanagi, T., Kawaguchi, H., & Kobayashi, T. (2012). Sequential memory scanning during sternberg memory task regardless of the different presentation style. *2012 ICME International Conference on Complex Medical Engineering (CME)*, 86–89. https://doi.org/10.1109/ICCME.2012.6275651

Ostaszewski, P., Green, L., & Myerson, J. (1998). Effects of inflation on the subjective value of delayed and probabilistic rewards. *Psychonomic Bulletin & Review*, *5*(2), 324–333. https://doi.org/10.3758/BF03212959

Piovesana, A., & Senior, G. (2018). How Small Is Big: Sample Size and Skewness. *Assessment*, *25*(6), 793–800. https://doi.org/10.1177/1073191116669784

Rabinowitz, A. R., Merritt, V. C., & Arnett, P. A. (2015). The Return-to-play incentive and the effect of motivation on neuropsychological test-performance: Implications for baseline concussion testing. *Developmental Neuropsychology*, *40*(1), 29–33. https://doi.org/10.1080/87565641.2014.1001066

Reddy, L. F., Horan, W. P., Barch, D. M., Buchanan, R. W., Dunayevich, E., Gold, J. M., Lyons, N., Marder, S. R., Treadway, M. T., Wynn, J. K., Young, J. W., & Green, M. F. (2015). Effort-based decision-making paradigms for clinical trials in schizophrenia: Part 1—Psychometric characteristics of 5 paradigms. *Schizophrenia Bulletin*, *41*(5), 1045–1054. psyh. https://doi.org/10.1093/schbul/sbv089

Renz, K. E., Pillny, M., & Lincoln, T. M. (2022). Increasing motivation in effort-based decision-making tasks: Effects of salience and reward expectancy manipulations. *Cognitive Neuropsychiatry*, *27*(1), 20–34. https://doi.org/10.1080/13546805.2021.2007068

Renz, K. E., Schlier, B., & Lincoln, T. M. (2023). Are effort-based decision-making tasks worth the effort?—A study on the associations between effort-based decision-making tasks and self-report measures. *International Journal of Methods in Psychiatric Research*, *32*(1), e1943. https://doi.org/10.1002/mpr.1943

Requena-Komuro, M.-C., Jiang, J., Dobson, L., Benhamou, E., Russell, L., Bond, R. L., Brotherhood, E. V., Greaves, C., Barker, S., Rohrer, J. D., Crutch, S. J., Warren, J. D., & Hardy, C. J. (2022). Remote versus face-to-face neuropsychological testing for dementia research: A comparative study in people with Alzheimer's disease, frontotemporal dementia and healthy older individuals. *BMJ Open*, *12*(11), e064576. https://doi.org/10.1136/bmjopen-2022-064576

Rickards, T. A., Cranston, C. C., Touradji, P., & Bechtold, K. T. (2018). Embedded performance validity testing in neuropsychological assessment: Potential clinical tools. *Applied Neuropsychology: Adult*, *25*(3), 219–230. https://doi.org/10.1080/23279095.2017.1278602

Robertson, J., & Kaptein, M. (2016). An Introduction to Modern Statistical Methods in HCI. In J. Robertson & M. Kaptein (Eds.), *Modern Statistical Methods for HCI* (pp. 1–14). Springer International Publishing. https://doi.org/10.1007/978-3-319-26633-6_1

Roebuck, H., Freigang, C., & Barry, J. G. (2016). Continuous performance tasks: Not just about sustaining attention. *Journal of Speech, Language, and Hearing Research : JSLHR*, *59*(3), 501–510. https://doi.org/10.1044/2015_JSLHR-L-15-0068

Rosenberg, J. M., Beymer, P. N., Anderson, D. J., Lissa, C. j van, & Schmidt, J. A. (2019). tidyLPA: An R Package to easily carry out latent profile analysis (LPA) using open-source or commercial software. *Journal of Open Source Software*, *3*(30), 978. https://doi.org/10.21105/joss.00978

Rosvold, H. E., Mirsky, A. F., Sarason, I., Bransome Jr., E. D., & Beck, L. H. (1956). A continuous performance test of brain damage. *Journal of Consulting Psychology*, *20*(5), 343–350. https://doi.org/10.1037/h0043220

Rydval, O., & Ortmann, A. (2004). How financial incentives and cognitive abilities affect task performance in laboratory settings: An illustration. *Economics Letters*, *85*(3), 315–320. https://doi.org/10.1016/j.econlet.2004.04.020

Salamone, J. D., Correa, M., Yang, J.-H., Rotolo, R., & Presby, R. (2018). Dopamine, Effort-Based Choice, and Behavioral Economics: Basic and Translational Research. *Frontiers in Behavioral Neuroscience*, *12*, 52. https://doi.org/10.3389/fnbeh.2018.00052

Saperia, S., Felsky, D., Da Silva, S., Siddiqui, I., Rector, N., Remington, G., Zakzanis, K. K., & Foussias, G. (2023). Modeling effort-based decision making: Individual differences in schizophrenia and major depressive disorder. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *8*(10), 1041–1049. https://doi.org/10.1016/j.bpsc.2023.05.009

Savulescu, J. (2001). The Fiction of "undue inducement": Why researchers should be allowed to pay participants any amount of money for any reasonable research project. *American Journal of Bioethics*. https://www.tandfonline.com/doi/abs/10.1162/152651601300169211

Schmidt, B., Hoffmann, E., & Rasch, B. (2020). Feel safe and money is less important! Hypnotic suggestions of safety decrease brain responses to monetary rewards in a risk game. *Cerebral Cortex Communications*, *1*(1), tgaa050. https://doi.org/10.1093/texcom/tgaa050

Schmidt, L., Lebreton, M., Cléry-Melin, M.-L., Daunizeau, J., & Pessiglione, M. (2012). Neural mechanisms underlying motivation of mental versus physical effort. *PLoS Biology*, *10*(2), e1001266. https://doi.org/10.1371/journal.pbio.1001266

Schroeder, R. W., Clark, H. A., & Martin, P. K. (2022). Base rates of invalidity when patients undergoing routine clinical evaluations have social security disability as an external incentive. *The Clinical Neuropsychologist*, *36*(7), 1902–1914. https://doi.org/10.1080/13854046.2021.1895322

Sheehan, D., Lecrubier, Y., Harnett Sheehan, K., Janavs, J., Weiller, E., Keskiner, A., Schinka, J., Knapp, E., Sheehan, M., & Dunbar, G. (1997). The validity of the Mini International Neuropsychiatric Interview (MINI) according to the SCID-P and its reliability. *European Psychiatry*, *12*(5), 232–241. https://doi.org/10.1016/S0924-9338(97)83297-X

Shum, D. H. K., O'Gorman, J. G., & Alpar, A. (2004). Effects of incentive and preparation time on performance and classification accuracy of standard and malingering-specific memory tests. *Archives of Clinical Neuropsychology*, *19*(6), 817–823. https://doi.org/10.1016/j.acn.2003.10.002

Slick, D. J., Sherman, E. M. S., & Iverson, G. L. (1999). Diagnostic Criteria for Malingered Neurocognitive Dysfunction: Proposed Standards for Clinical Practice and Research. *The Clinical Neuropsychologist*, *13*(4), 545–561. https://doi.org/10.1076/1385-4046(199911)13:04;1-Y;FT545

Soder, H. E., Cooper, J. A., Lopez-Gamundi, P., Hoots, J. K., Nunez, C., Lawlor, V. M., Lane, S. D., Treadway, M. T., & Wardle, M. C. (2021). Dose-response effects of d-amphetamine on effort-based decision-making and reinforcement learning. *Neuropsychopharmacology*, *46*(6), Article 6. https://doi.org/10.1038/s41386-020-0779-8

Sternberg, S. (1966). High-Speed Scanning in Human Memory. *Science*, *153*(3736), 652–654. https://doi.org/10.1126/science.153.3736.652

Stevens, A., Bahlo, S., Licha, C., Liske, B., & Vossler-Thies, E. (2016). Reaction time as an indicator of insufficient effort: Development and validation of an embedded performance validity parameter. *Psychiatry Research*, *245*, 74–82. https://doi.org/10.1016/j.psychres.2016.08.022

Strauss, G. P., Bartolomeo, L. A., & Luther, L. (2023). Reduced willingness to expend effort for rewards is associated with risk for conversion and negative symptom severity in youth at clinical high-risk for psychosis. *Psychological Medicine*, *53*(3), 714–721. https://doi.org/10.1017/S003329172100204X

Strauss, M. E., & Smith, G. T. (2009). Construct Validity: Advances in Theory and Methodology. *Annual Review of Clinical Psychology*, *5*, 1–25. https://doi.org/10.1146/annurev.clinpsy.032408.153639

Thomas, M. L., Patt, V. M., Bismark, A., Sprock, J., Tarasenko, M., Light, G. A., & Brown, G. G. (2017). Evidence of systematic attenuation in the measurement of cognitive deficits in schizophrenia. *Journal of Abnormal Psychology*, *126*(3), 312–324. https://doi.org/10.1037/abn0000256

Tombaugh, T. N. (1997). The Test of Memory Malingering (TOMM): Normative data from cognitively intact and cognitively impaired individuals. *Psychological Assessment*, *9*(3), 260–268. https://doi.org/10.1037/1040-3590.9.3.260

Tomczak, J., Gordon, A., Adams, J., Pickering, J. S., Hodges, N., & Evershed, J. K. (2023). What over 1,000,000 participants tell us about online research protocols. *Frontiers in Human Neuroscience*, *17*, 1228365. https://doi.org/10.3389/fnhum.2023.1228365

Tran, T. (2022). *Antecedents and Functional Outcomes of Motivation to Expend Physical and Cognitive Effort in Early Psychosis* [Doctor of Philosophy]. Queen's University.

Treadway, M. T., Buckholtz, J. W., Cowan, R. L., Woodward, N. D., Li, R., Ansari, M. S., Baldwin, R. M., Schwartzman, A. N., Kessler, R. M., & Zald, D. H. (2012). Dopaminergic Mechanisms of Individual Differences in Human Effort-Based Decision-Making. *Journal of Neuroscience*, *32*(18), Article 18. https://doi.org/10.1523/JNEUROSCI.6459-11.2012

Treadway, M. T., Buckholtz, J. W., Schwartzman, A. N., Lambert, W. E., & Zald, D. H. (2009). Worth the 'EEfRT'? The Effort Expenditure for Rewards Task as an objective measure of motivation and anhedonia. *PLOS ONE*, *4*(8), e6598. https://doi.org/10.1371/journal.pone.0006598

Treadway, M. T., Peterman, J. S., Zald, D. H., & Park, S. (2015). Impaired effort allocation in patients with schizophrenia. *Schizophrenia Research*, *161*(2–3), 382–385. psyh. https://doi.org/10.1016/j.schres.2014.11.024

Ulitzsch, E., Yildirim-Erbasli, S. N., Gorgun, G., & Bulut, O. (2022). An explanatory mixture IRT model for careless and insufficient effort responding in self-report measures. *British Journal of Mathematical and Statistical Psychology*, *75*(3), 668–698. https://doi.org/10.1111/bmsp.12272

van den Bos, W., Crone, E. A., & Güroğlu, B. (2012). Brain function during probabilistic learning in relation to IQ and level of education. *Developmental Cognitive Neuroscience*, *2*, S78–S89. https://doi.org/10.1016/j.dcn.2011.09.007

Van Patten, R. (2021). Introduction to the Special Issue - Neuropsychology from a distance: Psychometric properties and clinical utility of remote neurocognitive tests. *Journal of Clinical and Experimental Neuropsychology*, *43*(8), 767–773. https://doi.org/10.1080/13803395.2021.2021645

Vickery, C. D., Berry, D. T. R., Inman, T. H., Harris, M. J., & Orey, S. A. (2001). Detection of Inadequate Effort on Neuropsychological Testing: A Meta-Analytic Review of Selected Procedures. *Archives of Clinical Neuropsychology*, *16*(1), 45–73. https://doi.org/10.1093/arclin/16.1.45

Waltz, J. A., & Gold, J. M. (2016). Motivational Deficits in Schizophrenia and the Representation of Expected Value. In E. H. Simpson & P. D. Balsam (Eds.), *Behavioral Neuroscience of Motivation* (pp. 375–410). Springer International Publishing. https://doi.org/10.1007/7854_2015_385

Wang, Y., Wang, Y., Huang, J., Sun, X., Wang, X., Zhang, S., Zhu, G., Lui, S. S. Y., Cheung, E. F. C., Sun, H., & Chan, R. C. K. (2022). Shared and distinct reward neural mechanisms among patients with schizophrenia, major depressive disorder, and bipolar disorder: An effort-based functional imaging study. *European Archives of Psychiatry and Clinical Neuroscience*, *272*(5), 859–871. https://doi.org/10.1007/s00406-021-01376-3

Webb, S. S., Kontou, E., & Demeyere, N. (2022). The COVID-19 pandemic altered the modality, but not the frequency, of formal cognitive assessment. *Disability and Rehabilitation*, *44*(21), 6365–6373. https://doi.org/10.1080/09638288.2021.1963855

Wesnes, K. A., Brooker, H., Ballard, C., McCambridge, L., Stenton, R., & Corbett, A. (2017). Utility, reliability, sensitivity and validity of an online test system designed to monitor changes in cognitive function in clinical trials. *International Journal of Geriatric Psychiatry*, *32*(12), e83–e92. https://doi.org/10.1002/gps.4659

Whitely, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, *93*(1), 179–197. https://doi.org/10.1037/0033-2909.93.1.179

Wickens, C. D. (2014). Effort in Human Factors Performance and Decision Making. *Human Factors*, *56*(8), 1329–1336. https://doi.org/10.1177/0018720814558419

Wiley, K., Vedress, S., & Mandryk, R. L. (2020). How points and theme affect performance and experience in a gamified cognitive task. *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, 1–15. https://doi.org/10.1145/3313831.3376697

Yang, X., Huang, J., Zhu, C., Wang, Y., Cheung, E. F. C., Chan, R. C. K., & Xie, G. (2014). Motivational deficits in effort-based decision making in individuals with subsyndromal depression, first-episode and remitted depression patients. *Psychiatry Research*, *220*(3), 874–882. https://doi.org/10.1016/j.psychres.2014.08.056

Young, J. W., Light, G. A., Marston, H. M., Sharp, R., & Geyer, M. A. (2009). The 5-choice continuous performance test: Evidence for a translational test of vigilance for mice. *PLoS ONE*, *4*(1), e4227. https://doi.org/10.1371/journal.pone.0004227

Zedelius, C. M., Veling, H., Bijleveld, E., & Aarts, H. (2012). Promising High Monetary Rewards for Future Task Performance Increases Intermediate Task Performance. *PLoS ONE*, *7*(8), e42547. https://doi.org/10.1371/journal.pone.0042547

Zelazo, P. D., Anderson, J. E., Richler, J., Wallner-Allen, K., Beaumont, J. L., Conway, K. P., Gershon, R., & Weintraub, S. (2014). NIH Toolbox Cognition Battery (CB): Validation of executive function measures in adults. *Journal of the International Neuropsychological Society : JINS*, *20*(6), 620–629. https://doi.org/10.1017/S1355617714000472

Zou, Y.-M., Ni, K., Wang, Y.-Y., Yu, E.-Q., Lui, S. S. Y., Zhou, F.-C., Yang, H.-X., Cohen, A. S., Strauss, G. P., Cheung, E. F. C., & Chan, R. C. K. (2020). Effort–cost computation in a transdiagnostic psychiatric sample: Differences among patients with schizophrenia, bipolar disorder, and major depressive disorder. *PsyCh Journal*, *9*(2), 210–222. https://doi.org/10.1002/pchj.316