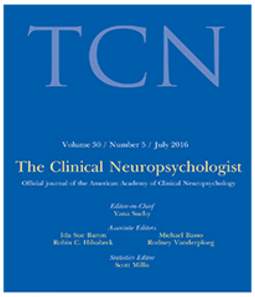

The Clinical Neuropsychologist

ISSN: 1385-4046 (Print) 1744-4144 (Online) Journal homepage: https://www.tandfonline.com/loi/ntcn20

The differential influence of computerized neuropsychological assessment across psychopathology

Shifali M. Singh, Jason R. Soble & Michael A. Young

To cite this article: Shifali M. Singh, Jason R. Soble & Michael A. Young (2019): The differential influence of computerized neuropsychological assessment across psychopathology, The Clinical Neuropsychologist, DOI: 10.1080/13854046.2019.1631888

To link to this article: https://doi.org/10.1080/13854046.2019.1631888

Published online: 26 Jun 2019.

THE CLINICAL NEUROPSYCHOLOGIST
https://doi.org/10.1080/13854046.2019.1631888

The differential influence of computerized neuropsychological assessment across psychopathology

Shifali M. Singh[a], Jason R. Soble[b,c] and Michael A. Young[a]

[a]Department of Psychology, Illinois Institute of Technology, Chicago, IL, USA; [b]Department of Psychiatry, University of Illinois College of Medicine, Chicago, IL, USA; [c]Department of Neurology, University of Illinois College of Medicine, Chicago, IL, USA

ABSTRACT

Objective: Novel technologies have transformed neuropsychological test administration so that research examining the equivalency of computerized versions of traditional tests is needed. This study examined the relationship between psychological symptom severity and performance on the Wisconsin Card Sorting Test (WCST) using the manual versus computerized administration.

Method: Eighty-five participants were randomly assigned to the manual or computerized WCST administration and also completed the Dot Counting Test, Depression Anxiety Stress Scales, and Short UPPS-P Impulsive Behavior Scale. Moderation analyses examined the effects of depressive, anxiety, and impulsivity symptom severity on WCST performance.

Results: For Perseverative Responses (PR), the methods of administration (MOAs) were equivalent regardless of psychological symptom severity. For failures to maintain set (FMS), MOA itself influenced performance, with participants making at least twice as many FMS on the computerized WCST. MOA also significantly moderated the relationship between FMS and impulsivity severity, including Lack of Perseverance, Sensation Seeking, and Positive Urgency. Individuals with greater Positive Urgency made more FMS on the manual WCST, and individuals with greater Lack of Perseverance made more FMS on the computerized WCST.

Conclusions: Findings suggest that equivalence between the manual and computerized versions of the WCST depends on the WCST subscale and the type of psychological symptom. New normative data need to be developed for the computerized WCST, along with a more consistent method of scoring and interpreting WCST subscales.

ARTICLE HISTORY
Received 23 March 2019
Accepted 9 June 2019
Publsihed online 26 June 2019

KEYWORDS
Neuropsychology; assessment; technology; executive functioning; psychopathology; computers

Introduction

The development, utilization, and efficacy of computers in psychological assessment are relatively new and are burgeoning areas of research. Improvements in digital accessibility have resulted in psychologists' incorporation of smartphones, the internet,

CONTACT Shifali M. Singh ssingh87@hawk.iit.edu Illinois Institute of Technology, Tech Central 201, 3424 S. State Street, Chicago, IL 60616

and software applications in research and practice (Fairburn & Patel, 2017; Moore et al., 2019). Computer-based neuropsychological assessment (CNA) is broadly defined as "any instrument that utilizes a computer, digital tablet, handheld device, or other digital interface instead of a human examiner to administer, score, or interpret tests of brain function and related factors relevant to questions of neurologic health and illness" (Bauer et al., 2012, p. 2). Such assessments have become increasingly popular in clinical practice over the last several decades (Lezak, 1983; Schatz, 2016).

There are several benefits to using CNAs. They ease the burden of the examiner while ensuring standardized administration, timing precision, and improved scoring accuracy (Paolo, Axelrod, Ryan, & Goldman, 1994; Woo, 2008; Neal & Grisso, 2014). Unlike human administrators, CNAs can quickly implement adaptive algorithms, which allow for complex manipulation during testing based on patients' real-time performance (Reise & Waller, 2009). Data from internet-based assessments can also contribute to large-scale databases that store patient information for research and medical purposes, and improve normative data collection (Bilder, 2011). However, concerns have been raised about factors such as limited patient-examiner interaction during administration, unfamiliarity with computers, and the psychometric properties of newly developed CNAs (Bauer et al., 2012; Fazeli, Ross, Vance, & Ball, 2013; Miller & Barr, 2017; Yantz & McCaffrey, 2007). Indeed, patients performed significantly worse when completing CNAs unobserved by examiners, compared to when they were observed (Yantz & McCaffrey, 2007). Tests that rely on reaction time or visual/spatial processing speed are more difficult for patients less familiar with using computers. Similarly, cognitively demanding CNAs present additional difficulty for individuals with computer inexperience, motor difficulties, or visual/hearing disabilities (Iverson, Brooks, Ashton, Johnson, & Gualtieri, 2009; Tanosoto et al., 2015).

Bauer et al. (2012) discussed problems programming traditional tests for computer administration, which has caused issues with the interpretability, validity, and reliability with psychometric data from CNAs. An example is the Wisconsin Card Sorting Test (WCST; Heaton et al., 1993). The WCST is a widely-used measure of cognitive flexibility with mixed findings regarding the equivalence between manual and computerized administrations (Artiola, Fortuny & Heaton, 1996; Feldstein et al., 1999; Heaton, Curtiss, & Tuttle, 1993; Hellman, Green, Kern, & Christenson, 1992; Steinmetz, Brunner, Loarer, Houssemand, 2010; Tien et al., 1996). Currently, both administrations of the WCST utilize the normative data from the manual administration. This is remarkable given the widespread use of the computerized WCST to assess executive functioning (EF), for which the validity and interpretability are essential. Finally, how psychopathology such as depression, anxiety, and impulsivity might interact with WCST method of administration (MOA) and affect performance and interpretation remains incompletely explored.

Depressive symptom severity is associated with deficits on tests of attention, memory, and EF, even controlling for slower processing speed or psychomotor symptoms associated with depression (Levin, Heller, Mohanty, Herrington, & Miller, 2007; Snyder, 2013). Administering CNAs requires additional considerations with individuals who have significant depressive symptom severity. For example, anhedonia is not only related to difficulty experiencing positive emotions or motivation, but also to deficits in reinforcement learning, incentive motivation, and reward-based decision-making (Pizzagalli, 2014; Watson et al., 1995). Motivation is necessary for successful completion of tests and the validity and interpretability of results (Heilbronner, Sweet, Morgan, Larrabee, & Millis, 2009). Treatment effectiveness studies show that MOA impacts motivation such that individuals are significantly more adherent and motivated to complete face-to-face compared to online treatment (Alfonsson, Johansson, Uddling, & Hursti, 2017). Self-Determination Theory states that motivation moderates one's willingness to successfully complete an assessment (Wenemark, Persson, Brage, Svensson, & Kristenson, 2011). Self-efficacy, the self-confidence a participant has to successfully accomplish a goal or task, also has been found to be lower in patients with depression, in part because it is related to lack of intrinsic motivation (Scheurich et al., 2008). Beyond the effects of depression on cognitive functioning, self-efficacy and motivation are particularly important for completing CNAs. Participants must complete tasks on the computer by themselves, with limited or no feedback from an examiner. Thus, individuals with depressive symptoms may perform more poorly on CNAs compared to alternative methods of assessment.

There are several potential mechanisms by which anxiety influences cognitive functioning. Anxiety severity is associated with deficits in attentional control (Eysenck et al., 2007), reaction times and working memory efficiency (Gustavson & Miyake, 2016), encoding and retrieving verbal episodic memory, and visual recognition memory (Delphin-Combe et al., 2016). If the MOA itself increases or decreases anxiety, it most likely will impact performance (Brünken & Plass, 2003; Buckelew & Hannay, 1986). Walther (2006) provided a theoretical framework for understanding how and why method of communication (online versus in-person) can influence individuals with anxiety. By eliminating the constant monitoring of nonverbal cues in face-to-face interaction, more cognitive resources can be allocated elsewhere. This is especially pertinent for neuropsychological tests that provide face-to-face verbal feedback, because negative verbal feedback during tests may be detrimental to performance due to increased anxiety. CNAs remove the potential for face-to-face negative feedback, which can be interpreted as threatening to individuals with symptoms of anxiety (High & Caplan, 2009). In line with this, when individuals with social anxiety disorder experience state emotion suppression after an anxiety-inducing social situation, they perform significantly more poorly than healthy controls on cognitive measures (O'Toole, Pedersen, Hougaard, & Rosenberg, 2015). Therefore, if interacting with an examiner induces anxiety, it may impact performance on cognitive tests. There are also implications for social desirability, in that individuals believe they represent their "true selves" more easily on computers, and endorse more psychological symptoms when answering personality questionnaires on a computer (Booth-Kewley, Larson, & Miyoshi, 2007). Stereotype threat related to gender and/or racial background may also impact test-takers, and lead to underperformance on cognitive measures in certain minority groups (Kapitanoff & Pandey, 2017; Thames et al., 2013). Thus, individuals with anxiety may perform better on CNAs versus alternative methods given their hypervigilance to nonverbal cues, beliefs about social desirability, emotion suppression, and potential factors related to stereotype threat.

Impulsivity is a multifaceted construct that consists of motor (action without thinking), attentional (lack of focus on the task at hand), and planning (orientation towards the present, rather than to the future) components (Barratt, 1959). Significant impulsivity is associated with attentional deficits, behavioral disinhibition, and executive dysfunction (Hervey, Epstein, & Curry, 2004; Pennington & Ozonoff, 1996). Many CNAs have already been developed to assess impulsivity because computerized testing provides more real-time data, such as measures of reaction time, which is relevant to impulse-related disorders (Riccio, Reynolds, Lowe, & Moore, 2002). CNAs are able to identify impulsivity more accurately in those who may attempt suicide than self-report questionnaires typically used in conjunction with CNAs in neuropsychology outpatient clinics (Horesh, 2001). In attention-based neurological conditions, CNAs are more successful in identifying neglect than traditional paper-and-pencil tests (Bonato, Priftis, Umilta, & Zorzi, 2013). Impulsivity is also highly correlated with drug addiction, and computer-based interventions used in the assessment and treatment of drug addiction are effective in increasing motivation, length of sobriety, and access to treatment (Bickel, Christensen, & Marsch, 2011). Motivation is particularly important for patients with elevated impulsivity, because motivation and self-efficacy are often reduced (Adler, 2010). CNAs may be more successful in assessing cognitive functioning for individuals who are socially disinhibited because they reduce the amount of social interaction (Kertzman, Reznik, Grinspan, & Shliapnicov, 2006). The literature suggests individuals with impulsivity may perform better on CNAs versus alternative methods, due to improved motivation when using computerized interfaces, greater difficulty sustaining attention on paper-and-pencil tasks, and increased social disinhibition.

The present study

Although depression, anxiety, and impulsivity have been shown to influence performance on cognitive tasks in general, and particularly on measures of executive functioning, it is unclear how that influence may differ by MOA. This study assessed how MOA and psychological symptom severity separately and jointly influence performance on the WCST. By understanding how symptoms contribute to performance based on MOA, clinicians will have more certainty when identifying patients who would benefit from a CNA versus manual administration. This has implications for the future of neuropsychology, as perhaps novel CNAs can be developed in a way that makes computerized testing preferred for individuals with a wide range of psychopathology. The following were hypothesized.

Hypothesis 1: MOA moderates the relationship between depressive symptom severity and WCST performance, such that those with greater depressive symptoms will perform better on the manual version versus the computerized administration.

Hypothesis 2: MOA moderates the relationship between anxiety symptom severity and WCST performance, such that those with greater anxiety symptom severity will perform better on the computerized compared to manual administration.

Hypothesis 3: MOA moderates the relationship between impulsivity symptom severity and WCST performance, such that those with greater impulsivity symptom severity will perform better on the computerized administration relative to the manual version.

Table 1. Participant characteristics.

	Manual ($n = 44$)	Computerized ($n = 41$)	Total ($n = 85$)
Age, mean years (SD)	34.7 (15.4)	33.8 (16.4)	34.3 (1.7)
Female, *n* (%)	26 (59.1)	21 (51.2)	47 (55.3)
Handedness, *n* (%)			
Right	40 (90.9)	37 (90.2)	77 (90.6)
Left	2 (4.5)	3 (7.3)	5 (5.9)
Both	2 (4.5)	1 (2.4)	3 (3.5)
Education, mean years (SD)	13.2 (1.8)	12.9 (2.1)	13.0 (1.9)
Race/Ethnicity, *n* (%)			
White	28 (63.6)	14 (34.1)	42 (49.4)
Hispanic/Latino	3 (6.8)	4 (9.8)	7 (8.2)
Black/Af. Am.	6 (13.6)	17 (41.5)	23 (27.1)
Native Am.	1 (2.3)	–	1 (1.2)
Asian	1 (2.3)	4 (9.8)	5 (5.9)
Other	5 (11.4)	2 (4.9)	7 (8.2)
Computer use, *n* (%)			
$>1\times$/day	25 (56.8)	26 (63.4)	51 (60.0)
Smartphone use, *n* (%)			
$>1\times$/day	38 (86.4)	35 (85.4)	73 (85.9)
Participant type, *n* (%)			
Outpatient	22 (50.0)	20 (48.8)	42 (49.4)
Student	15 (34.1)	16 (39.0)	31 (35.6)
Public	7 (15.9)	5 (12.2)	12 (14.1)
Medication use, *n* (%)	22 (50.0)	13 (31.7)	35 (41.1)

Method

Participants

This cross-sectional study utilized data collected from 85 participants recruited from outpatient mental health facilities ($n = 42$), university students ($n = 31$), and the general public ($n = 12$) for a study of cognitive and emotional functioning in which they would be playing a card game. Inclusion criteria consisted of self-reports of English as one's first language, being born in the United States, and no history of neurological illness. Recruitment methods aimed to collect data from individuals with a broad range of depressive, anxiety, and impulsivity symptom severity. Participants were recruited using flyers that asked whether they struggled with "emotional issues," "sadness or hopelessness," "tension or anxiety," "smoking or drug addiction," and whether they "regularly gamble." Upon recruitment, participants were randomly assigned to a testing condition. Every participant successfully passed a performance validity test (described below), which ensured valid neuropsychological test performance. Participant characteristics are presented in Table 1.

The manual and computerized MOA groups did not significantly differ by age ($F(1,83) = .08$, $p = .78$), gender ($X^2(1, N = 85) = .532$, $p = .47$), handedness (X^2 (1, $N = 85$) $= .545$, $p = .76$), education ($F(1,83) = 1.35$, $p = .25$), participant type (X^2 (2, $N = 85$) $= .355$, $p = 0.84$), or current use of psychiatric medications (yes or no; X^2 (1, $N = 85$) $= 2.932$, $p = .09$). The majority of participants reported using computers (60.0%) and smartphones (85.9%) "More than once a day." Race/ethnicity was dichotomized into a "White" and "Non-White" group, and significantly differed between manual and computerized groups (X^2 (1, $N = 85$) $= 7.384$, $p = .01$). To determine whether race/ethnicity significantly differed across other demographic variables, ANOVAs were computed for race/ethnicity and age, gender, handedness, education,

Table 2. Descriptive statistics for Perseverative Responses, failures to maintain set, depression, anxiety, and impulsivity by method of administration.

	Manual mean (SD)	Computerized mean (SD)	*p* value	Skew statistic
Perseverative Responses	49.3 (13.79)	53.07 (13.61)	.21	.21
Failures to maintain set	.48 (.73)	1.17 (1.69)	.02	2.54
Depressive severity	9.89 (9.53)	9.73 (11.92)	.95	.50
Anxiety severity	9.00 (8.18)	10.44 (11.59)	.51	.50
Impulsivity				
Negative urgency	1.49 (1.05)	1.23 (1.17)	.28	.40
Lack of Perseverance	2.00 (1.00)	2.03 (1.18)	.90	.44
Lack of premeditation	1.47 (.98)	1.42 (.95)	.82	.46
Sensation Seeking	0.95 (1.03)	.99 (1.18)	.85	−0.22
Positive Urgency	1.33 (.73)	1.22 (.82)	.51	0.25

$n = 85$.

participant type, use of psychiatric medications, computer use, and smartphone use, and were all nonsignificant (*ps* = .23–.55). In addition, using ANOVA and Poisson Regression, race/ethnicity was not significantly related to PR ($F(1,83) = .511$, $p = .48$) or FMS ($ß = -.670$, 95% CI = [−1.359, .019], $p = .06$). Therefore, race/ethnicity was not treated as a covariate in analyses.

There were no significant differences between the manual and computerized administration groups on the DASS depression ($F(1,83) = .004$, $p = .95$) or anxiety ($F(1,83) = .442$, $p = .51$) subscales. Likewise, for impulsivity, there were no significant differences between the manual and computerized groups on the Negative Urgency ($F(1,83) = 1.174$, $p = .28$), Lack of Perseverance ($F(1,83) = .016$ $p = .90$), Lack of Premeditation ($F(1,83) = .055$, $p = .82$), Sensation Seeking ($F(1,83) = .039$, $p = .845$), and Positive Urgency ($F(1,83) = .429$, $p = .51$) impulsivity scales. Group differences by MOA are presented in Table 2.

Measures

The dot counting test (DCT; Boone, Lu, & Herzberg, 2002). This test was designed to detect invalid neuropsychological test performance. An "*E*-Score" of ≥14 was used to identify invalid performance, which is associated with 88% sensitivity and 96% specificity.

Depression anxiety stress scales (DASS; Lovibond & Lovibond, 1995). The DASS is a 42-item self-report questionnaire that measures dimensions of depression, anxiety, and stress/tension. Each item contributes to only one scale. Participants circle a number from "0" to "3" describing how much each statement applied to them over the past week. Some of the items associated with depression include: feeling downhearted/blue, feeling life is meaningless or not worthwhile, and feeling worthless. Some anxiety-related items include: feeling shaky, feeling worry about situations, and being scared for no good reason. Stress-related items include difficulty winding down, overreacting, and feeling a state of nervous tension. The depression and anxiety scale scores used in analyses for this study were computed by summing the scores of items within each scale to obtain a raw score.

Short UPPS-P Impulsive Behavior Scale (SUPPS-P; Whiteside & Lynam, 2001). The SUPPS-P is a 45-item scale developed based on factor analysis of 8 commonly used measures of impulsivity, and measures four facets of impulsivity: (1) Urgency, (2) Lack of Premeditation, (3) Lack of Perseverance, and (4) Sensation Seeking. Lynam, Smith, Whiteside, and Cyders (2006) split the "Urgency" component into both "Positive Urgency" and "Negative Urgency," creating the 5-point scale, 59-item UPPS-P. The original UPPS only accounted for a "tendency to commit rash and regrettable actions as a result of intense *negative* affect" and not "positive affect." This study used the well-validated short-form of the UPPS-P, or the SUPPS-P (Lynam, 2013), which consists of 20 items on a 4-point scale (Agree Strongly, Agree Some, Disagree Some, or Strongly Disagree), with 4 items for each of the 5 subscales. The SUPPS-P was scored using manual norms, and higher scores indicate higher levels of impulsivity. Each of the 5 subscales were examined separately.

WCST (Heaton, Chelune, Talley, Kay, & Curtiss, 1993). The WCST is a widely-used test of cognitive flexibility that measures abstraction, conceptual ability, and ability to shift one's attention, and has both a manual and computerized version (Harris, 1990; Heaton, Curtiss, & Tuttle, 1993). A large sample study ($N = 1221$) found a three-factor structure, including cognitive flexibility, problem-solving, and response maintenance (Greve, Stickle, Love, Bianchini, & Stanford, 2005). The three-factor model indicators were (1) Perseverative Responses, Conceptual Level Responses, Categories Completed, and Total Correct; (2) Conceptual Level Responses, Categories Completed, Total Correct, and Nonperseverative Errors; and (3) Total Correct and Failures to Maintain Set. Perseverative Responses (PR) are responses in which one continues to sort a correct sorting rule from the previous stage, and meets the conditions enumerated in the 1981 manual. PR is considered the variable most sensitive to deficits related to psychopathology, and is the most widely used variable in research studying cognitive flexibility using this task (Heaton, 1981). The PR score also accounts for a change in perseverative principle, wherein the participant "makes three unambiguous incorrect matches in succession according to another principle" (Heaton, 1981; p. 22).

Failures to maintain set (FMS) is when a participant is given positive feedback for five consecutive cards, but is unable to maintain the rule and reach the criterion for completing the category. FMS has been related to both executive and attentional functioning (Greve, Ingram, & Bianchini, 1998). Given the unique indicators that PR and FMS measure (i.e., cognitive flexibility and response maintenance), both PR and FMS were used as the primary measures of WCST performance. Because of the common usage of demographically adjusted T-scores for age and education, and eliminating the need to use age and education as covariates in analyses, this study used demographically adjusted T-scores for PR. Because FMS did not follow a normal distribution, raw scores were used instead. No variables from the second factor were included given the overlap of several categories between the first and second factors, and the high correlation between PR and nonperseverative error *T*-scores ($r = .79$, $p < 0.001$).

Participants were administered either the computerized or manual version of the standard, 128-card version of WCST (Heaton, Chelune, Talley, Kay, & Curtiss, 1993; Heaton et al., 1993). During the test, they matched response cards to key cards using novel matching rules that they were not aware of beforehand, and received feedback indicating whether they were correct or incorrect after each response.

The computerized version is similar to the manual version, except for (1) the absence of an examiner; and (2) "right" or "wrong" feedback provided by a voice from the computer with accompanying written text displayed on the screen.

Self-report questionnaire

Participants reported age, gender, handedness, years of education, race/ethnicity, computer use, smartphone use, and medication use for psychiatric reasons (yes or no). They rated whether their computer and smartphone use in the past 2 weeks was "never," "less than once a week," "once a week," "2–3 times a week," "4–6 times a week," "once a day," or "more than once a day."

Procedure

This study was approved by the IRB. All subjects provided written and informed consent. Data were collected in a university, outpatient setting, and the Chicagoland community. Due to funding limitations, data was collected from 85 participants. Participants were assigned a random identification number to ensure confidentiality. These numbers were randomized before collecting data and were assigned to participants as they were recruited. Once recruited, participants were asked to indicate that English was their first language, that they were between ages 18 and 69, that they were born in the United States, and that they did not have any history of a neurological disorder. Participants with even identification numbers were assigned to the manual version of the WCST; those with odd numbers were assigned to the computerized version. The order of the test battery was the WCST (manual or computerized version), the DCT, the DASS, and the SUPPS-P, and took between 45 and 60 minutes to complete. One examiner administered all assessments in this study, and was present in the room for the entirety of the assessments, including the computerized WCST. The WCST was administered first to eliminate any influence from the computer administration of other tasks. The same computer was used for all participants when administering the WCST to eliminate the chance of different computer settings (i.e., brightness, screen size, trackpad sensitivity) variably impacting performance. Following the DCT, participants were provided with a laptop with a unique Qualtrics link and completed demographic questions. After demographic questions, they completed the DASS, SUPPS-P, and answered a question about medication usage. Instructions for the DASS and SUPPS-P were read aloud to ensure participants understood the tasks and relative time periods of questions. Responses were uploaded to the Qualtrics database immediately after completion. Following testing, participants were compensated with a $15 gift card for their time. Students were given the option of a $15 gift card or 1 extra credit point.

Data analyses

Data were analyzed using SPSS version 24 (IBM, 2017). Moderation analyses using the SPSS PROCESS macro (Hayes, 2013) were computed separately for each symptom severity type (depressive, anxiety, or impulsivity) to determine whether MOA (manual versus computerized) moderated the relationship between symptom severity and either PR or FMS. Covariates included depended on whether there was a significant relationship between a variable (i.e., age, gender, handedness, education, race/ethnicity, computer use, smartphone use, and medication use) and either PR or FMS. For all moderation analyses, if the interaction effect was not significant, the regression analysis was rerun omitting the interaction term to assess the main effects.

Results

Preliminary data analyses

The majority of participants reported using computers (60.0%) and smartphones (85.9%) "More than once a day," so the "computer use" and "smartphone use" groups were dichotomized into "Frequent" and "Infrequent" groups, with "More than once a day" in the "Frequent" group, and all other responses in the "Infrequent" group (Table 2). Participant type was dichotomized into "clinical" and "nonclinical" groups, with the "clinical" group including outpatients, and the "nonclinical" group including students and those recruited from the general public. After dichotomizing, manual and computerized administration groups did not significantly differ based on computer use (X^2 (1, $N = 85$) $=$ .385, p $=$ .54), smartphone use (X^2 (1, $N = 85$) $=$ .017, $p = .90$), or participant type (X^2 (1, $N = 85$) $=$.223, p $=$ .64). Data were normally distributed for all measures. No measure was significantly skewed or kurtotic ($p < .05$; Table 2).

Tests of hypotheses for PR

Descriptive statistics for WCST PR, depressive symptom severity, anxiety symptom severity, and impulsivity symptom severity are presented in Table 2. Performances on the WCST PR varied widely among participants and were not significantly different between administration groups (Table 2). The PR *T*-scores on the WCST are already adjusted for age and education, so neither of those variables were included in covariate analyses. Handedness was not considered a covariate because 90.6% participants were right-handed. One-way ANOVAs were computed between PR and all other potential covariates: gender, computer use, smartphone use, and psychiatric medication. PR was significantly related to smartphone use (F(1,83) $=$ 0.04), so it was used as a covariate in moderation analyses. Moderation analyses were conducted to determine whether MOA moderated the relationship between psychological symptom severity and PR on the WCST (Figure 1). All symptom severity variables were centered prior to analyses. Smartphone use was used as a covariate in each analysis. In summary, none of the interaction or main effects were significant (Table 3); all main effects remained nonsignificant when the interaction effect was omitted.

Tests of hypotheses for FMS

Descriptive statistics for WCST FMS and severities of depressive, anxiety, and impulsivity symptoms are presented in Table 2. As noted earlier, there were no significant

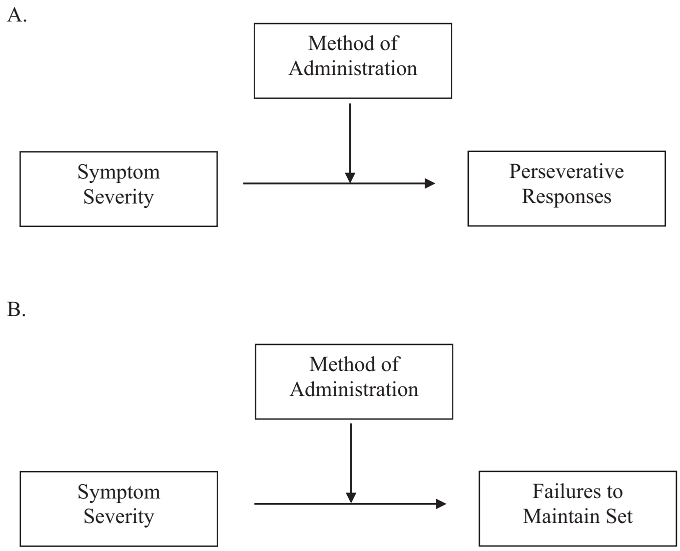

Figure 1. Moderation model for symptom severity, method of administration, and Perseverative Responses.

differences between the manual and computerized administration groups on severity of depression, anxiety, or impulsivity (see Table 2) and data for these variables were approximately normally distributed. The FMS subscale was not normally distributed and followed a Poisson-like distribution. This distribution might have been expected because FMS is a count of the number of failures to maintain set. Therefore, regression analyses used a Poisson model or a negative binomial model if the distribution was over-dispersed (Coxe, West, & Aiken, 2009). Individual regressions were computed for the associations between FMS and the potential covariates of age, gender, education, computer use, smartphone use, and psychiatric medication use. FMS was significantly related only to age ($ß = .03$, 95% CI = [.011, .045], $p = .001$] so it was used as a covariate in FMS moderation analyses to determine whether MOA moderated the relationship between psychological symptom severity and FMS (Figure 1). All symptom severity variables were centered prior to analyses.

Results for MOA moderating the effect of the five subscales of impulsivity on FMS were mixed (Table 4). Consistent with hypotheses, Lack of Perseverance, Sensation Seeking, and Positive Urgency had significant interactions, whereas Negative Urgency and Lack of Premeditation did not.

The interaction of Lack of Perseverance and MOA was significant. With the computerized version, greater Lack of Perseverance resulted in a significantly greater number of failures to maintain set ($ß = .388$, 95% CI = [.113, .662], $p = .01$). An increase of one point in Lack of Perseverance was associated with 1.47 times the number of failures to maintain set. This was not the case for the manual administration ($ß = -.182$, 95% CI = [−.670, .306], $p = .47$).

The interaction of Sensation Seeking and MOA was significant. The main effect of Sensation Seeking was not significant in either the manual ($ß = .356$, 95% CI = [−.080, .792], $p = .11$) or the computerized administration groups ($ß = -.244$, 95% CI = [−.495,

Table 3. Regression of method of administration moderating the effect of symptom severity on Perseverative Responses.

Depression (*D*)[a]	*B*	*SE*	*t*	*p*
Smartphone	−1.896	1.041	−1.821	.072
D	.083	.231	.358	.721
MOA (M)	3.146	3.294	.955	.343
D × M interaction	−.033	.289	−.114	.909
Anxiety (A)[b]				
Smartphone	−2.038*	1.010	−2.018	.047
A	.234	.261	.898	.372
MOA (M)	2.987	2.974	1.005	.318
A × M interaction	−.096	.317	−.304	.762
Negative urgency (N)[c]				
Smartphone	−1.684	.986	−1.708	.092
N	.628	1.973	.318	.751
MOA (M)	3.207	2.985	1.074	.286
N × M interaction	−2.280	2.757	−.827	.411
Lack of Perseverance (LP)[d]				
Smartphone	−1.952	.997	−1.959	.054
LP	.869	2.112	.411	.682
MOA (M)	3.313	2.926	1.132	.261
LP × M interaction	−4.049	2.801	−1.445	.152
Lack of premeditation (LPr)[e]				
Smartphone	−2.014*	.974	−2.068	.042
LPr	2.321	2.107	1.101	.274
MOA (M)	3.218	2.911	1.106	.272
LPr × M interaction	−6.102	3.076	−1.984	.051
Sensation Seeking (S)[f]				
Smartphone	−1.444	1.018	−1.418	.160
S	1.931	2.081	.928	.356
MOA (M)	3.340	2.961	1.128	.263
S × M interaction	−.772	2.734	−.282	.778
Positive Urgency (P)[g]				
Smartphone	−1.703	.989	−1.723	.089
P	−.668	2.854	−.234	.816
MOA (M)	3.239	2.985	1.085	.281
P × M interaction	−.528	3.903	−.135	.893

Smartphone = Smartphone Use, MOA = Method of Administration (Manual versus Computerized).
[a] R^2 = .059, F(4, 80) = 1.253, p = .30.
[b] R^2 = .072, F(4, 80) = 1.552, p = .195.
[c] R^2 = .066, F(4,80) = 1.421, p = .235.
[d] R^2 = .093, F(4, 80) = 2.060, p = .094.
[e] R^2 =.102, F(4, 80) = 2.281, p = .068.
[f] R^2 = .071, F(4, 80) = 1.529, p = .202.
[g] R^2 = .060 F(4, 80) =1.269, p = .289.
n = 85.

.006], p = .06). The opposite directions of these two effects probably accounts for the interaction indicating their difference, despite neither themselves effect being significantly different from zero.

The interaction of Positive Urgency and MOA was significant. In the manual administration, greater Positive Urgency resulted in significantly more failures to maintain set (*ß* = .821, 95% CI = [.299, 1.343], p = .002). An increase of one point in of Positive Urgency was associated with 2.27 times the number of failures to maintain set. This was not the case for the computerized administration (*ß* = −.024, 95% CI = [−.500, .452], p = .92).

The interaction of negative urgency and MOA was not significant; however, the main effect of MOA on FMS was significant (*ß* =.940, 95% CI = [.418, 1.462], p < .001),

Table 4. Poisson/negative binomial regressions of method of administration moderating the effect of symptom severity on failures to maintain set.

Depression (*D*)	*ß*	*SE*	*95% CI*	*p*	*Exp(B)*
Age	.033***	.011	[.011, .054]	.003	1.0134
MOA (M)	−.914*	.400	[−1.698, −.131]	.022	.401
D	−.003	.019	[−.040, .035]	.895	.998
D x M interaction	−.021	.038	[−.095, .053]	.577	.979
Anxiety (A)					
Age	.032**	.011	[.010, .054]	.004	1.032
MOA (M)	−.816*	.350	[−1.502, −.129]	.020	.442
A	−.002	.020	[−.042, .037]	.908	.998
A x M interaction	.001	.040	[−.077, .079]	.983	1.001
Negative urgency (N)					
Age	.031***	.008	[.016, .045]	<.001	1.031
MOA (M)	−1.022***	.288	[−1.586, −.458]	<.001	.360
N	−.026	.126	[−.274, .221]	.835	.974
N x M interaction	.342	.231	[−.111, .794]	.139	1.408
Lack of Perseverance (LP)					
Intercept	−1.211***	.358	[−1.912, −.510]	.001	.298
Age	.033***	.0073	[.018, .047]	<.001	1.033
MOA (M)	−.807***	.273	[−1.342, −.271]	.003	.446
LP	.381***	.116	[.154, .609]	.001	1.464
LP x M interaction	−.606*	.271	[−1.137, −.075]	.025	.546
Lack of premeditation (LPr)					
Intercept	−1.087***	.3492	[−1.771, −.402]	.002	.337
Age	.032***	.007	[.017, .046]	<.001	1.032
MOA (M)	−.877***	.265	[−1.400, −.357]	.001	.416
LPr	.254	.131	[−.002, .510]	.052	1.289
LPr x M interaction	−.188	.252	[−.681, .305]	.455	.829
Sensation Seeking (S)					
Intercept	−1.141***	.358	[−1.842, −.440]	.001	.319
Age	.032***	.008	[.017, .047]	<.001	1.033
MOA (M)	−.855***	.273	[−1.390, −.319]	.002	.425
S	−.249	.128	[−.499, .000]	.050	.779
S x M interaction	.646***	.248	[.161, 1.132]	.009	1.908
Positive Urgency (P)					
Intercept	−.992***	.344	[−1.665, −.318]	.004	.371
Age	.030***	.008	[.015, .044]	<.001	1.030
MOA (M)	−1.114***	.302	[−1.705, −.523]	<.001	.328
P	.736**	.2676	[.211, 1.260]	.006	2.087
P x M interaction	.811*	.328	[.167, 1.454]	.014	2.250

MOA = Method of Administration (0 = Manual, 1 = Computerized).
*$p < .05$.
**$p < .01$.
***$p < .001$.
$n = 85$.

Compared to manual administration, computer administration was associated with 2.56 times the number of failures to maintain set. The main effect of Negative Urgency was not significant (*ß* = .075, 95% CI = [-.126, .276] $p = .47$).

The interaction of lack of premeditation and MOA not significant. The main effect MOA was significant (*ß* =.916, 95% CI = [.403, 1.429], $p < .001$). Compared to manual version, computer administration was associated with 2.50 times the number of failures to maintain set. The main effect of Lack of Premeditation was not significant (*ß* =.201, 95% CI = [−.018, .420], $p = .07$).

MOA did not significantly moderate the relationship between FMS and depressive or anxiety symptom severity. However, the main effect of MOA on FMS was significant in the analyses for both symptoms (depression: *ß* = .811, 95% CI = [.124, 1.498],

$p = .02$, anxiety: $ß = .815$, 95% CI = [.129, 1.502], $p = .02$). Compared to manual version, computer administration was associated with 2.25 and 2.26 times the number of failures to maintain set, respectively.

In summary, there were no significant interaction or main effects regarding the effects of MOA and psychological symptom severity on PR. However, there were some significant moderating and main effects of MOA on FMS, in particular when examining impulsivity. In addition, regardless of psychological symptom severity, there were at least twice as many FMS with the computerized administration of the WCST relative to the manual administration.

Discussion

This study examined whether the MOA of the WCST influenced the relationship between psychopathology and performance. Although other studies have separately investigated (1) the influence of psychopathology on WCST performance (Iznak et al., 2016; Keilp et al., 2018); (2) differences in performances between the computerized and manual administration of the WCST (Artiola, Fortuny & Heaton, 2007; Feldstein et al., 1999; Tien et al., 1996); (3) the relation between MOA and psychopathology (Ozonoff, 1995); and (4) the influence of CNAs on performance in general (Hämmäinen, 1994; Yantz & McCaffrey, 2007), this is the first study to examine the relationship among all these concepts together. Understanding how MOA influences performance is especially pertinent in this current era of neuropsychology, during which novel CNAs are replacing traditional MOAs (Schatz, 2016).

Several studies have attempted to address equivalence between traditionally paper-and-pencil tests, and computerized versions of traditionally paper-and-pencil tests, including the Tower of London, Halstead Category Test, and the WCST (Artiola, Fortuny & Heaton, 1996; Hellman et al., 1992; Nici and Hom, 2013; Williams & Jarrold, 2013). These studies all concluded that performance on tests were not significantly different based on MOA. In contrast to these findings, and those demonstrating equivalence of subscales of the WCST with different MOAs (Tien et al., 1996), our findings suggest that the equivalence of the manual and computerized versions of the WCST depends on the WCST subscale and the type of psychological symptom studied. In line with Steinmetz et al. (2010), participants made at least twice as many failures to maintain set with computerized administration. A greater Lack of Perseverance was associated with greater FMS in the computerized administration, and a greater Positive Urgency was associated with a greater FMS in the manual administration. These findings corroborate studies utilizing FMS to specifically measure distractibility and impulsivity in participants with impulse-related disorders (e.g., Kaplan, Şengör, Gürvit, Genç, & Güzeliş, 2006).

Lack of Perseverance is defined in the SUPPS-P as an inability to remain with a task until completion (Whiteside & Lynam, 2001). It follows that those with greater Lack of Perseverance have greater FMS, but this effect was only observed in the computerized administration of the WCST. This may be because during the manual administration, participants are able to move cards as quickly as they choose by placing the card under the choice card, but during the computerized administration, once a participant chooses a place for a card to move, they must wait for the card for a predetermined amount of time to reach its destination. If participants had difficulty sustaining attention because of the relatively long time it takes for a card to move to a chosen location, which is implicated in studies on impulsivity (e.g., Hervey, Epstein, & Curry, 2004), it might at least partially explain why these findings were only observed in the computerized administration of the WCST. Lack of Perseverance is also defined as the tendency to not continue an activity perceived as unengaging (Mallorquí-Bagué et al., 2018). It is conceivable that participants felt more engaged in the manual administration because they were working with a human examiner (Yantz & McCaffrey, 2007).

Greater Positive Urgency resulted in significantly more FMS in the manual administration, but was not significantly associated with FMS in the computerized administration. Although these findings are seemingly the opposite of the results with Lack of Perseverance, they can be explained by examining impulsive traits and WCST MOAs. Positive Urgency describes a tendency towards a "rash" action in response to a very positive mood, and FMS are defined as a set-loss after five consecutive correct responses (Yuen & Lee, 2003). After receiving positive feedback after five consecutive correct responses, the elevated mood associated with that positive feedback may have contributed to participants experiencing Positive Urgency, and therefore incurring greater FMS.

It is likely for the same reasons that participants with greater Lack of Perseverance made more FMS on the computerized administration, participants with greater Positive Urgency made more FMS on the *manual* administration. That is, although participants with greater Positive Urgency tended to make more FMS when given continuous positive feedback and the ability to move cards as quickly or slowly as desired, the computerized administration forced participants to slowly and consistently match each card.

Clinicians should exercise caution when interpreting the WCST for patients who may have greater Lack of Perseverance and Positive Urgency, which have been associated with inattention, aggression, gambling use disorder, borderline personality disorder, antisocial personality disorder, attention-deficit hyperactivity disorder (in children), and alcohol use disorder (Cyders et al., 2007; Mallorquí-Bagué et al., 2018; Miller, Derefinko, Lynam, Milich, & Fillmore, 2010; Whiteside, Lynam, Miller, & Reynolds, 2005). Additionally, the discrepancy in FMS between MOAs illustrates the necessity for normative data that can equate the manual and computerized versions of the WCST. Clinicians should conservatively interpret data from the computerized WCST, because it is conceivable that other WCST subscales not analyzed might also differ based solely on MOA.

This study included features in its design unlike previous studies comparing WCST MOAs (i.e., Feldstein et al., 1999; Steinmetz et al., 2010; Tien et al., 1996). It utilized a relatively ethnically and diagnostically heterogeneous sample, with nearly half of participants taking medication for psychiatric reasons and/or recruited from an outpatient mental health facility. Unlike previous studies (Steinmetz et al., 2010; Tien et al., 1996), participants in this study were given either the manual or the computerized administration, not both, so that experience with one type of administration could not affect performance with the other type. Participants were also asked how familiar they were with technology, which was considered in analyses. This study utilized a performance validity measure to ensure that performances on the WCST were both effortful and accurate assessments of cognitive functioning. Findings can be immediately utilized in clinical practice, and can provide insight for future test developers and researchers using the WCST.

This study also contributes to the broader EF literature. EF is a construct that reflects both states and traits, and is significantly influenced by contextual factors such as pain, sleep quality, and stress (Holanda Junior & Almondes, 2016; Krabbe, Ellbin, Nilsson, Jonsdottir, & Samuelsson, 2017; Murata et al., 2017). This study demonstrates that psychiatric symptom severity may significantly and differentially impact test performance, particularly for those with high Lack of Perseverance and Positive Urgency. Given that EF significantly impacts performance-based instrumental activities of daily living (IADLs), these findings have implications for not only EF as measured by WCST performance, but also for predicting psychiatrically-related change in successful completion of IADLs in daily life (Bell-McGinty, Podell, Franzen, Baird, & Williams, 2002; Jefferson, Paul, Ozonoff, & Cohen, 2006; Vaughn & Giovanello, 2010). Therefore, although participants may score within normal limits on neuropsychological tests, they may have psychiatrically related difficulties completing IADLs in daily life.

Though there was a wide range of reported races/ethnicities, there were an insufficient number of participants from each racial/ethnic category to conduct analyses based on distinct racial/ethnic groups. Race/ethnicity was dichotomized into a “White” and “Non-White” group for analyses, which is a limitation of this study in that several different racial/ethnic groups (e.g., Native American, Asian American) were combined into one group despite their group differences. Further, race/ethnicity significantly differed between manual and computerized groups, though not significantly on PR or FMS. Future studies should therefore aim to collect data from a significant number of participants in each racial/ethnic category such that factors related to race/ethnicity can be considered in analyses. Additionally, these results can only be generalized to individuals between the ages of 18–66. Older adults were excluded to eliminate chances for comorbid medical/health or neurologic diagnoses, and to reduce any effects of technology familiarity. Future research might examine how MOA influences individuals above age 69, particularly as researchers develop new methods for diagnosing dementias and other age-related neurodegenerative disorders. Although all participants spoke English as their first language and were born in the United States, there was no assessment to rule out whether a participant had difficulty reading, which could have led to misinterpretation of questions.

Another potential limitation is a lack of sufficient power to reduce the chance of Type-II error. Findings suggest depression and anxiety do not influence performance on either of the WCST subscales studied, and that only two scales of impulsivity influence performance on the FMS subscale. It is possible that a larger sample size would better identify any influence of psychopathology on WCST subscale performance.

Although the most commonly used WCST subscales were investigated, there is no standard for determining which of the WCST subscales should be used for interpreting performance. Some suggest PR is the best indicator of frontal lobe dysfunction (Heaton, 1981), and others have utilized different subscales, such as FMS (Kaplan, Şengör, Gürvit, Genç, & Güzeliş, 2006), non-perseverative errors (Barceló, 2001), or total errors (Keilp et al., 2018). In addition, there are no available normative data specific to the computerized WCST. Once these norms are established, this research should be replicated to determine whether participants are still significantly more likely to have FMS in the computerized WCST, and whether there are significant differences in performance based on psychopathology and WCST subscale. Once the relation between performance and psychopathology is better understood, normative data can be established that can help clinicians more accurately determine whether an aberrant performance was due to true neuropsychological deficits, or to an artifact of psychopathology. It might be useful to update the computerized WCST to more closely mimic paper-and-pencil administration conditions, perhaps by using concepts from Item Response Theory (Reise & Waller, 2009; Thomas, 2011), wherein the computer interface is adaptive and can match the speed of the examinee.

As neuropsychology transitions towards utilizing computerized neuropsychological assessments more frequently, it will become increasingly important to continue evaluating psychometric properties of these assessments. The significance of establishing normative data for computerized assessments cannot be emphasized more given these findings, which suggest that performance on neuropsychological assessments can be significantly different based on MOA. Once these new norms are established, researchers can better delineate the differential influence of psychopathology on WCST performance, so that practitioners are able to make clinical decisions based on more accurate assessment data.

References

Adler, L. A. (2010). Monitoring adults with ADHD: A focus on executive and behavioral function. *The Journal of Clinical Psychiatry*, *71*(8), 18.

Alfonsson, S., Johansson, K., Uddling, J., & Hursti, T. (2017). Differences in motivation and adherence to a prescribed assignment after face-to-face and online psychoeducation: An experimental study. *BMC Psychology*, *5*(1), 3.

Artiola I., Fortuny, L. A., & Heaton, R. K. (1996). Standard versus computerized administration of the Wisconsin card sorting test. *The Clinical Neuropsychologist*, *10*(4), 419–425.

Artiola, I., Fortuny, L., & Heaton, R. K. (2007). The clinical neuropsychologist standard versuscomputerized administration of the Wisconsin Card Sorting Test. *The Clinical Neuropsychologist*, *10*(4), 419–424. doi:10.1080/13854049608406702

Barceló, F. (2001). Does the Wisconsin card sorting test measure prefontral function? *The Spanish Journal of Psychology*, *4*(1), 79–100.

Barratt, E. S. (1959). Anxiety and impulsiveness related to psychomotor efficiency. *Perceptual and Motor Skills*, *9*(3), 191–198.

Bauer, R. M., Iverson, G. L., Cernich, A. N., Binder, L. M., Ruff, R. M., & Naugle, R. I. (2012). Computerized neuropsychological assessment devices: Joint position paper of the American Academy of Clinical Neuropsychology and the National Academy of Neuropsychology. *The Clinical Neuropsychologist*, *26*(2), 177–196.

Bell-McGinty, S. M., Podell, K. P., Franzen, M., Baird, A. D., & Williams, M. J. (2002). Standard measures of executive function in predicting instrumental activities of daily living in older adults. *International Journal of Geriatric Psychiatry*, *17*(9), 828–834.

Bickel, W. K., Christensen, D. R., & Marsch, L. A. (2011). A review of computer-based interventions used in the assessment, treatment, and research of drug addiction. *Substance Use & Misuse*, *46*(1), 4–9. doi:10.3109/10826084.2011.521066

Bilder, R. M. (2011). Neuropsychology 3.0: Evidence-based science and practice. *Journal of the International Neuropsychological Society*, *17*(01), 7–13.

Bonato, M., Priftis, K., Umilta, C., & Zorzi, M. (2013). Computer-based attention-demanding testing uneils severe neglect in apparently intact patients. *Behavioral Neurology*, *26*(3), 179–181.

Boone, K., Lu, P., & Herzberg, D. (2002). *The dot counting test*. Los Angeles: Western Psychological Services.

Booth-Kewley, S., Larson, G. E., & Miyoshi, D. K. (2007). Social desirability effects on computerized and paper-and-pencil questionnaires. *Computers in Human Behavior. Computers in Human Behavior*, *23*(1), 463–477. doi:10.1016/j.chb.2004.10.020

Brünken, R., & Plass, J. L. (2003). Direct measurement of cognitive load in multimedia. *Learning*, *38*, 53–61.

Buckelew, S., & Hannay, J. (1986). Relationships among anxiety, defensiveness, sex, task difficulty, and performance on various neuropsychological tasks. *Perceptual and Motor Skills*, *63*(2), 711–718.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. *Statistical Development and Applications*, *91*(2), 121–136.

Cyders, M. A., Smith, G. T., Spillane, N. S., Fischer, S., Annus, A. M., & Peterson, C. (2007). Integration of impulsivity and positive mood to predict risky behavior: Development and validation of a measure of positive urgency. *Psychological Assessment 19*(1), 107–118.

Delphin-Combe, F., Bathsavanis, A., Rouch, I., Liles, T., Vannier-Nitenberg, C., Fantino, B., … Krolak-Salmon, P. (2016). Relationship between anxiety and cognitive performance in an elderly population with a cognitive complaint. *European Journal of Neurology*, *23*(7), 1210–1217.

Eysenck, M. W., Derakshan, N., Santos, R., & Calvo, M. G. (2007). Anxiety and cognitive performance: Attentional control theory. *Emotion*, *7*(2), 336. doi:10.1037/1528-3542.7.2.336

Fairburn, C. G., & Patel, V. (2017). The impact of digital technology on psychological treatments and their dissemination. *Behaviour Research and Therapy*, *88*, 19–25.

Fazeli, P. L., Ross, L. A., Vance, D. E., & Ball, K. (2013). The relationship between computer experience and computerized cognitive test performance among older adults. *Journals of Gerontology Series B: Psychological Sciences and Social Sciences*, *68*(3), 337–346.

Feldstein, S. N., Keller, F. R., Portman, R. E., Durham, R. L., Klebe, K. J., & Davis, H. P. (1999). A comparison of computerized and standard versions of the Wisconsin card sorting test. *The Clinical Neuropsychologist*, *13*(3), 303–313. doi:10.1076/clin.13.3.303.1744

Greve, K. W., Ingram, F., & Bianchini, K. J. (1998). Latent structure of the Wisconsin card sorting test in a clinical sample. *Archives of Clinical Neuropsychology*, *13*(7), 597–609. doi:10.1016/S0887-6177(97)00075-9

Greve, K. W., Stickle, T. R., Love, J. M., Bianchini, K. J., & Stanford, M. S. (2005). Latent structure of the Wisconsin card sorting test: A confirmatory factor analytic study. *Archives of Clinical Neuropsychology*, *20*(3), 355–364.

Gustavson, D. E., & Miyake, A. (2016). Trait worry is associated with difficulties in working memory updating. *Cognition and Emotion*, *30*(7), 1289–1303.

Hämmäinen, L. (1994). Computerized support for neuropsychological test interpretation in clinical situations. *Clinical Neuropsychologist*, *8*(2), 167–185. doi:10.1080/13854049408401555

Harris, M. E. (1990). *Wisconsin card sorting test: Computerized version, research edition*. Odessa, FL: Psychological Assessment Resources, Inc.

Hayes, A. F. (2013). Model templates for PROCESS for SPSS and SAS. Retrieved from http://www.afhayes.com/public/templates.pdf.

Heaton, R. K. (1981). *A manual for the Wisconsin card sorting test*. Odessa, FL: Psychological Assessment Resources, Inc.

Heaton, R. K., Chelune, G. J., Talley, J. L., Kay, G. G., & Curtiss, G. (1993). *Wisconsin card sorting test (WCST) manual revised and expanded.* Odessa, FL: Psychological Assessment Resources, Inc.

Heaton, R. K., Curtiss, G., & Tuttle, K. (1993). *Wisconsin card sorting test: Computer version –2.* Odessa, FL: Psychological Assessment Resources, Inc.

Heilbronner, R. L., Sweet, J. J., Morgan, J. E., Larrabee, G. J., & Millis, S. R. (2009). American academy of clinical neuropsychology consensus conference statement on the neuropsychological assessment of effort, response bias, and malingering. *Clinical Neuropsychology*, *23*(7), 1093–1129.

Hellman, S. G., Green, M. F., Kern, R. S., & Christenson, C. D. (1992). Comparison of card and computer versions of the Wisconsin Card Sorting Test for psychotic patients. *International Journal of Methods in Psychiatric Research*, *2*, 151–155.

Hervey, A. S., Epstein, J. N., & Curry, J. F. (2004). Neuropsychology of adults with attention-deficit/hyperactivity disorder: A meta-analytic review. *Neuropsychology*, *18*(3), 485.

High, A. C., & Caplan, S. E. (2009). Social anxiety and computer-mediated communication during initial interactions: Implications for the hyperpersonal perspective. *Computers in Human Behavior*, *25*(2), 475–482.

Holanda Junior, F. W. N., & Almondes, K. M. D. (2016). Sleep and executive functions in older adults: A systematic review. *Dementia and Neuropsychologia*, *10*(3), 185–197.

Horesh, N. (2001). Self-report vs. computerized measures of impulsivity as a correlate of suicidal behavior. *The Journal of Crisis Intervention and Suicide Prevention*, *22*(1), 27–31.

Iverson, G. L., Brooks, B. L., Ashton, V. L., Johnson, L. G., & Gualtieri, C. T. (2009). Does familiarity with computers affect computerized neuropsychological test performance? *Journal of Clinical and Experimental Neuropsychology*, *31*(5), 594–604.

Iznak, A. F., Medvedeva, T. I., Iznak, E. V., Oleichik, I. V., Bologov, P. V., & Kobzova, M. P. (2016). Original Russian text. *Human Physiology*, *42*(6), 18–26.

Jefferson, A. L., Paul, R. H., Ozonoff, A., & Cohen, R. A. (2006). Evaluating elements of executive functioning as predictors of instrumental activities of daily living. *Archives of Clinical Neuropsychology*, *21*(4), 311–320.

Kapitanoff, S., & Pandey, C. (2017). Stereotype threat, anxiety, instructor gender, and underperformance in women. *Active Learning in Higher Education*, *18*(3), 213–229. doi:10.1177/1469787417715202

Kaplan, G. B., Şengör, N. S., Gürvit, H., Genç, İ., & Güzeliş, C. (2006). A composite neural network model for perseveration and distractibility in the Wisconsin card sorting test. *Neural Networks*, *19*(4), 375–387. doi:10.1016/j.neunet.2005.08.015

Keilp, J. G., Madden, S. P., Gorlyn, M., Burke, A. K., Oquendo, M. A., & Mann, J. J. (2018). The lack of meaningful association between depression severity measures and neurocognitive performance. *Journal of Affective Disorders*, *241*, 164–172. doi:10.1016/j.jad.2018.08.034

Kertzman, S., Reznik, I., Grinspan, H., & Shliapnicov, N. (2006). The role of real-time computerized neuropsychological examination in forensic psychiatry practice. *The Israel Journal of Psychiatry and Related Sciences*, *43*(3), 174.

Krabbe, D., Ellbin, S., Nilsson, M., Jonsdottir, I. H., & Samuelsson, H. (2017). Executive function and attention in patients with stress-related exhaustion: Perceived fatigue and effect of distraction. *Stress*, *20*(4), 333–340.

Levin, R. L., Heller, W., Mohanty, A., Herrington, J. D., & Miller, G. A. (2007). Cognitive deficits in depression and functional specificity of regional brain activity. *Cognitive Therapy and Research*, *31*(2), 211–233. doi:10.1007/s10608-007-9128-z

Lezak, M. D. (1983). *Neuropsychological assessment* (2nd ed.). London, UK: Oxford Press.

Lovibond, S. H., & Lovibond, P. F. (1995). *Manual for the depression anxiety stress scales.* Sydney: Psychology Foundation.

Lynam, D. R. (2013). Development of a short form of the UPPS-P impulsive behavior scale. (Technical Report). West Lafayette: Purdue University.

Lynam, D. R., Smith, G. T., Whiteside, S. P., & Cyders, M. A. (2006). *Assessing five personality pathways to impulsive behavior* (Technical report). West Lafayette: Purdue University.

Mallorquí-Bagué, N., Mestre-Bach, G., Lozano-Madrid, M., Fernandez-Aranda, F., Granero, R., Vintró-Alcazaz, C., … Jiménez-Murcia, S. (2018). Trait impulsivity and cognitive domains involving impulsivity and compulsivity as predictors of gambling disorder treatment response. *Addictive Behaviors*, *87*, 169–176. doi:10.1016/j.addbeh.2018.07.006

Miller, J. B., & Barr, W. B. (2017). The technology crisis in neuropsychology. *Archives in Clinical Neuropsychology*, *32*(5), 541–554.

Miller, D. J., Derefinko, K. J., Lynam, D. R., Milich, R., & Fillmore, M. T. (2010). Impulsivity and attention deficit-hyperactivity disorder: Subtype classification using the UPPS impulsive behavior scale. *Journal of Psychopathology and Behavioral Assessment*, *32*(3), 323–332. doi:10.1007/s10862-009-9155-z

Moore, T. M., Gur, R. C., Thomas, M. L., Brown, G. G., Nock, M. K., Savitt, A. P., … Stein, M. B, … & Army STARRS Collaborators. (2019). Development, administration, and structural validity of a brief, computerized neurocognitive battery: Results from the army study to assess risk and resilience in servicemembers. *Assessment*, *26*(1), 125–143. doi:10.1177/1073191116689820

Murata, S., Sawa, R., Nakatsu, N., Saito, T., Sugimoto, T., Nakamura, R., … Ono, R. (2017). Association between chronic musculoskeletal pain and executive function in community-dwelling older adults. *European Journal of Pain*, *21*(10), 1717–1722. doi:10.1002/ejp.1083

Neal, T., & Grisso, T. (2014). Assessment practices and expert judgment methods in forensic psychology and psychiatry: An international snapshot. *Criminal Justice and Behavior*, *41*(12), 1406–1421.

Nici, J., & Hom, J. (2013). Comparability of the computerized Halstead Category Test with the original version. *Archives of Clinical Neuropsychology*, *28*(8), 824–828.

O'Toole, M. S., Pedersen, A. D., Hougaard, E., & Rosenberg, N. K. (2015). Neuropsychological test performance in social anxiety disorder. *Nordic Journal of Psychiatry*, *69*(6), 1726–1734.

Ozonoff, S. (1995). Reliability and validity of the Wisconsin card sorting test in studies of autism. *Neuropsychology*, *9*(4), 491–500.

Paolo, A. M., Axelrod, B. N., Ryan, J. J., & Goldman, R. S. (1994). Administration accuracy of the Wisconsin card sorting test. *The Clinical Neuropsychologist*, *8*(1), 112–116.

Pennington, B. F., & Ozonoff, S. (1996). Executive functions and developmental psychopathology. *Journal of Child Psychology and Psychiatry*, *37*(1), 51–87.

Pizzagalli, D. A. (2014). Depression, stress, and anhedonia: Toward a synthesis and integrated model. *Annual Review of Clinical Psychology*, *10*(1), 393–423.

Reise, S. P., & Waller, N. G. (2009). Item response theory and clinical measurement. *Annual Review of Clinical Psychology*, *5*(1), 27–48.

Riccio, C. A., Reynolds, C. R., Lowe, P., & Moore, J. J. (2002). The continuous performance test: A window on the neural substrates for attention? *Archives of Clinical Neuropsychology*, *17*(3), 235–272.

Schatz, P. (2016). Computer-based assessment: Current status and next steps. In R. L. Kane, & T. D. Parsons (Eds.), *The role of technology in clinical neuropsychology* (pp. 47–60). New York, NY: Oxford University Press.

Scheurich, A., Fellgiebel, A., Schermuly, I., Bauer, S., Wölfges, R., & Müller, M. J. (2008). Experimental evidence for a motivational origin of cognitive impairment in major depression. *Psychological Medicine*, *38*(2), 237–246.

Snyder, H. R. (2013). Major depressive disorder is associated with broad impairments on neuropsychological measures of executive function: A meta-analysis and review. *Psychological Bulletin*, *139*(1), 81.

Steinmetz, J. P., Brunner, M., Loarer, E., & Houssemand, C. (2010). Incomplete psychometric equivalence of scores obtained on the manual and the computer version of the Wisconsin card sorting test? *Psychological Assessment*, *22*(1), 199.

Tanosoto, T., Bendixen, K. H., Arima, T., Hansen, J., Terkelsen, A. J., & Svensson, P. (2015). Effects of the paced auditory serial addition task (PASAT) with different rates on autonomic nervous system responses and self-reported levels of stress. *Journal of Oral Rehabilitation*, *42*(5), 378–385.

Thames, A. D., Hinkin, C. H., Byrd, D. A., Bilder, R. M., Duff, K. J., Rivera Mindt, M., … Streiff, V. (2013). Effects of stereotype threat, perceived discrimination, and examiner race on neuropsychological performance: Simple as Black and White? *Journal of the International Neuropsychological Society*, *19*(5), 583–593. doi:10.1017/S1355617713000076

Thomas, M. L. (2011). The value of item response theory in clinical assessment: A review. *Assessment*, *18*(3), 291–307.

Tien, A. Y., Spevack, T. V., Jones, D. W., Pearlson, G. D., Schlaepfer, T. E., & Strauss, M. E. (1996). Computerized Wisconsin card sorting test: Comparison with manual administration. *The Kaohsiung Journal of Medical Sciences*, *12*(8), 479–485.

Vaughn, L., & Giovanello, K. (2010). Executive function in daily life: Age-related influences of executive processes on instrumental activities of daily living. *Psychology and Aging*, *25*(2), 343–355. doi:10.1037/a0017729

Walther, J. B. (2006). *Nonverbal dynamics in computer-mediated communication, or:(and the net:('s with you,:) and you:) alone.* Thousand Oaks, CA: Sage Publications.

Watson, D., Weber, K., Assenheimer, J. S., Clark, L. A., Strauss, M. E., & McCormick, R. A. (1995). Testing a tripartite model: I. Evaluating the convergent and discriminant validity of anxiety and depression symptom scales. *Journal of Abnormal Psychology*, *104*(1), 3–14.

Wenemark, M., Persson, A., Brage, H., Svensson, T., & Kristenson, M. (2011). Applying motivation theory to achieve increased respondent satisfaction, response rate and data quality in a self-administered survey. *Journal of Official Statistics*, *27*(2), 393–414.

Whiteside, S. P., & Lynam, D. R. (2001). The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689. doi:10.1016/S0191-8869(00)00064-7

Whiteside, S. P., Lynam, D. R., Miller, J. D., & Reynolds, S. K. (2005). Validation of the UPPS impulsive behaviour scale: A four-factor model of impulsivity. *European Journal of Personality* , *19*(7), 559–574. doi:10.1002/per.556

Williams, D., & Jarrold, C. (2013). Assessing planning and set-shifting abilities in Autism: Are experimenter and computerized versions of tasks equivalent? *Autism Research*, *6*(6), 461–467. doi:10.1002/aur.1311

Woo, E. (2008). Computerized neuropsychological assessments. *CNS Spectrums*, *13*(S16), 14–17. doi:10.1017/S1092852900026985

Yantz, C. J., & McCaffrey, R. J. (2007). Social facilitation effect of examiner attention or inattention to computer-administered neuropsychological tests: First sign that the examiner may affect results. *The Clinical Neuropsychologist*, *21*(4), 663–671. doi:10.1080/13854040600788158

Yuen, K. S. L., & Lee, T. M. C. (2003). Could mood state affect risk-taking decisions? *Journal of Affective Disorders*, *75*(1), 11–18. doi:10.1016/S0165-0327(02)00022-8