UC Riverside
UC Riverside Previously Published Works

Title

UCancellation: A new mobile measure of selective attention and concentration

Permalink

https://escholarship.org/uc/item/9zx0h0jk

Journal

Behavior Research Methods, 54(5)

ISSN

1554-351X

Authors

Pahor, Anja
Mester, Randy E
Carrillo, Audrey A
et al.

Publication Date

2022-10-01

DOI

10.3758/s13428-021-01765-5

Peer reviewed

UCancellation: A new mobile measure of selective attention and concentration

Anja Pahor[1,2] · Randy E. Mester[3] · Audrey A. Carrillo[3] · Eunice Ghil[1] · Jason F. Reimer[4] · Susanne M. Jaeggi[5] · Aaron R. Seitz[3,2]

Abstract

Measuring selective attention in a speeded task can provide valuable insight into the concentration ability of an individual, and can inform neuropsychological assessment of attention in aging, traumatic brain injury, and in various psychiatric disorders. There are only a few tools to measure selective attention that are freely available, psychometrically validated, and can be used flexibly both for in-person and remote assessment. To address this gap, we developed a self-administrable, mobile-based test called "UCancellation" (University of California Cancellation), which was designed to assess selective attention and concentration and has two stimulus sets: Letters and Pictures. UCancellation takes less than 7 minutes to complete, is automatically scored, has multiple forms to allow repeated testing, and is compatible with a variety of iOS and Android devices. Here we report the results of a study that examined parallel-test reliability and convergent validity of UCancellation in a sample of 104 college students. UCancellation Letters and Pictures showed adequate parallel test reliability ($r = .71–.83$, $p < 0.01$) and internal consistency ($\alpha = .73–.91$). It also showed convergent validity with another widely used cancellation task, d2 Test of Attention ($r = .43–.59$, $p < 0.01$), and predicted performance on a cognitive control composite ($r = .34–.41$, $p < 0.05$). These results suggest that UCancellation is a valid test of selective attention and inhibitory control, which warrants further data collection to establish norms.

Introduction

Selective attention refers to the ability to restrict behavioral responses to a relevant set of stimuli, while concurrently ignoring irrelevant information. It is fundamental to successful performance on most cognitive tasks, and deficits in

✉ Anja Pahor
apahor@uci.edu

1   School of Education, University of California, Irvine, Irvine, CA, USA

2   Department of Psychology, University of California, Riverside, Riverside, CA, USA

3   Brain Game Center for Mental Fitness and Wellbeing, University of California, Riverside, Riverside, CA, USA

4   Department of Psychology, California State University, San Bernardino, San Bernardino, CA, USA

5   School of Education, School of Social Sciences (Department of Cognitive Sciences), University of California, Irvine, Irvine, CA, USA

these processes have broad implications for development, and academic achievement (Brodeur & Pond, 2001; Jonkman, 2005; Stevens & Bavelier, 2012). Selective attention difficulties have been reported for individuals with acquired brain injury (Spaccavento et al., 2019; Ziino & Ponsford, 2006), neurodegenerative disorders (Duchek et al., 2009; Levinoff et al., 2004; Maddox et al., 1996), and even in normal aging (Commodari & Guarnera, 2008; McDowd & Filion, 1992; Zanto & Gazzaley, 2004).

Cancellation tasks involve searching for and crossing out targets that are embedded among distractor stimuli and are as such often used to assess selective and sustained attention, as well as psychomotor speed and response inhibition. While initially developed to measure spatial neglect in individuals with acquired brain injury (Ferber & Karnath, 2001), the use of cancellation tasks has expanded to other patient groups and to healthy individuals (Dalmaijer et al., 2015; Della Sala et al., 1992; Huang & Wang, 2009; Uttl & Pilkenton-Taylor, 2001). For example, one of the supplemental subtests of the Wechsler Adult Intelligence Scale IV is Cancellation,

which is intended for use in adults aged 16–70 years, and its score contributes to a Processing Speed Index. Perhaps one of the most widely used cancellation tasks is d2 Test of Attention (Brickenkamp & Zillmer, 1998), which involves crossing out certain letters but not others, and has been validated extensively in European and US populations (Bates & Lemay, 2004). Blotenberg and Schmidt-Atzert (2019) postulated that performance on cancellation tasks involves four sub-components: perception of an item, performing a simple mental operation on the item, a motor reaction (e.g., select, cross-out), and shifting to the next item. In a series of experiments, the authors demonstrated that the first two sub-components (perceptual and mental operation speed) are the strongest predictors of performance on sustained attention tasks. Further, the postulated sub-components did not explain a large amount of variance in working memory span and reasoning tasks, thereby providing evidence of discriminant validity of cancellation tasks.

Access to reliable and valid online tests of cognitive abilities is becoming increasingly important for continuity of research and clinical assessment in light of restricted in-person testing during the COVID-19 pandemic. Although there are a few free tools for computerized administration of cancellation tasks (Dalmaijer et al., 2015; Rorden & Karnath, 2010; Wang et al., 2006), these are not necessarily optimized for online research or do not have cross-platform availability. With these constraints in mind, we developed a mobile device-based test of selective attention and concentration called *UCancellation (University of California Cancellation)*. The test involves responding to multiple rows of items and selecting as many targets as possible within a time limit. There are two versions that can be used interchangeably: *Letters* and *Pictures*. UCancellation Letters is similar to the d2 Test of Attention (Brickenkamp & Zillmer, 1998) in that it consists of similar stimuli: characters "d" and "p" with one to four marks, arranged either individually or in pairs above or below the letters. In UCancellation Pictures, letters are replaced with pictures of animals, some of which are targets while others are distractors.

UCancellation emulates standard cancellation tasks, which involve crossing out target shapes (e.g. stars, bells), numbers, or letters arranged in either random or structured arrays of items (Weintraub and Mesulam, 1985; Halligan et al., 1989; Gauthier et al., 1989; Ferber & Karnath, 2001), as well as the d2 (Brickenkamp & Zillmer, 1998), a widely used paper and pencil test of concentration performance that builds upon preceding cancellation tests. The d2 test can be used for individuals aged 9–60 years and consists of a single form, which can be administered individually or in group format. The d2 test consists of 14 lines, each comprised of 47 characters "d" and "p" with one to four marks, arranged either individually or in pairs above or below the character. The participant is asked to scan the lines from left to right and cross out all "d"s with two marks within a time limit of 20 seconds per line. There are extensive norms based on clinical and nonclinical populations from Europe (Brickenkamp & Zillmer, 1998; Rivera et al., 2017) that have been extended to populations in other regions. Several studies have demonstrated that the d2 is highly reliable (Brickenkamp & Zillmer, 1998; Lee et al., 2018; Steinborn et al., 2018) and that it shows good construct validity (Bates & Lemay, 2004; Wassenberg et al., 2008). The d2 manual, 20 recording blanks, and a set of two scoring keys can be purchased at Hogrefe Publishing Corp for $137 excluding tax. Additional recording blanks need to be acquired to accommodate larger sample sizes. Hogrefe also offers an online version of the d2 test of Attention–Revised (d2-R), which consists of 14 screens of 60 symbols displayed in six rows of 10, and can be accessed for approximately the same price as the paper and pencil version. While both the paper and the computerized d2-R version continue to be of value to researchers, there is a need for high-quality, freely accessible tests that can be administered on participants' own devices, with data outputs that are automatically scored and easy to interpret, which was the goal of UCancellation.

Here we report the results of a study that examined the parallel-test reliability and convergent validity of UCancellation Letters and UCancellation Pictures using an undergraduate student population. Convergent validity was investigated by comparing tablet-based UCancellation performance with paper and pencil d2. We predicted that both versions of UCancellation would be positively correlated with d2 performance. We further investigated how both of these measures relate to a Cognitive Control (CC) composite derived from select computerized tasks from the NIH EXAMINER (Kramer et al., 2014), and another tablet-based inhibitory control task developed in our lab (Countermanding). We predicted that both versions of UCancellation would be positively correlated with the Cognitive Control composite and negatively correlated with Countermanding reaction time measures (i.e. better performance on UCancellation would be related to faster performance on Countermanding). Since UCancellation Pictures requires switching between two target types whereas UCancellation Letters only uses one target, it was hypothesized that performance on the former would be more strongly related to an index of shifting on the Countermanding task. In addition, we explored whether performance on the UCancellation task would be associated with self-report measures of personality, cognitive failures, hours spent video game playing, and motivation for goal-directed activities, as suggested by previous research (Green & Bavelier, 2003; Prabhakaran et al., 2011; Robertson et al., 1997; van Steenbergen et al., 2009).

Table 1 Demographics for 104 participants presented separately for each group

	Pictures group	Letters group
Age		
Mean (standard deviation)	20.8 years (4.5)	21.6 years (6.0)
Gender		
Female	66.0%	76.0%
Male	32.1%	22.0%
"Unknown" or "Do not want to specify"	1.9%	2.0%
Race		
American Indian/Alaska Native	1.9%	0.0%
Asian	39.6%	40.0%
Black or African American	1.9%	4.0%
Native Hawaiian or Other Pacific Islander	1.9%	0.0%
White	22.6%	14.0%
More than one ethnicity/race	11.3%	6.0%
"Unknown" or "Do not want to specify"	20.8%	36.0%
Ethnicity		
Hispanic or Latino	39.6%	44.0%
Not Hispanic or Latino	60.4%	54.0%
"Unknown" or "Do not want to specify"	0.0%	2.0%

Note. Questions regarding race and ethnicity and the resulting data are presented separately as defined by the US Office of Management and Budget Standards

Methods

Participants

118 undergraduate students from the University of California, Irvine and University of California, Riverside participated in the study, 60 of which were randomly assigned to complete UCancellation Pictures and 58 to complete UCancellation Letters. The participants provided informed consent, received course credit or $10 an hour for participation, and had normal or corrected-to-normal vision. Fourteen participants were excluded due to incomplete session data or corrupted data in the Countermanding task, an issue that has been resolved since, yielding a final sample size of 104 participants ($N_{Letters}$ = 50, $N_{Pictures}$ = 54). The demographics of the two groups of participants are presented in Table 1.

Procedure

Participants were randomly assigned to two groups: one group completed UCancellation Letters in sessions 1 and 2, while the other completed UCancellation Pictures in both sessions. The sessions took place 1 week apart at the same time of day and at the same laboratory. During these sessions, both groups also completed a set of surveys, a Countermanding task, the paper-pencil d2, and four computerized cognitive control measures: Flanker, Set-Shifting, Continuous Performance Task and Antisaccades from NIH EXAMINER. The order of tasks was counterbalanced across participants to control for order effects. Specifically, d2 was performed in the same session as the surveys and the Countermanding task, while the NIH EXAMINER was performed in the other session (the order of sessions was also counterbalanced).

Measures

UCancellation

UCancellation is a self-administrable, mobile device-based assessment of selective attention and inhibitory control. The application was developed in the Unity game engine and is currently supported on iOS and Android (note that it can also be run on desktop computers; however, the validation to date refers to the use on touch screens). Stimuli consist of either letters "d" and "p" or pictures of dogs and monkeys arranged in rows (cf. Fig. 1). A tutorial explains how to perform the task, followed by three practice rows. A smiley face is displayed after practice if there were at least two "perfect rows" (all targets selected, no errors); conversely, a sad face is displayed if there were fewer than two perfect rows and, in this scenario, practice is repeated to ensure that the participant understands the task. In the test phase, the rows are displayed one at a time in the center of the screen for 6 seconds, with a 1-second blank screen interval between rows, and no feedback is presented. Each row
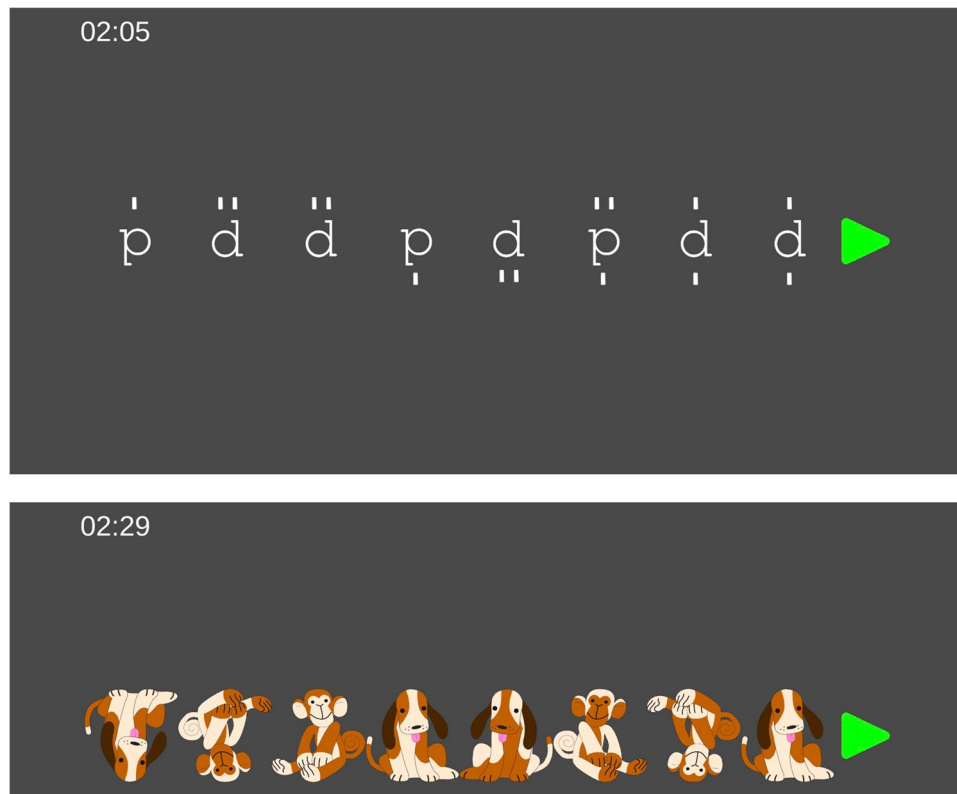
Fig. 1 Screenshots of UCancellation Letters (top) and Pictures (bottom). A countdown timer is present in the top left corner of the screen and a green arrow can be pressed to advance to the next line (if row time has not been exceeded)

consists of eight stimuli, 3–5 of which are targets, aiming for approximately 40 targets in 10 rows[1]. Each row is populated with 3–5 nontargets (lures) using random sampling without replacement. Retaking the test will result in a different sequence of items within the constraints mentioned above. The goal is to select all targets from left to right within a 6-second time limit. An auditory cue signals that time is out, and a different sound cue is presented if the participant achieves a "perfect row." If a participant clears a row before the time limit, they can press a button to continue to the next row. Thus, some participants can complete bonus rows as long as a global time limit of 3 minutes and 30 seconds is not exceeded. A countdown timer indicating the global time limit is presented in the top left corner of the screen. At the end of the task, the participant gets feedback on how many perfect rows they cleared.

Data text files are stored on the device itself and can also be shared on an Amazon Web Services HIPAA-compliant[2]

server. There are three types of data files: Row, Action, and Summary. Row data files contain the following information per row: number of hits, misses, false alarms and correct rejections, last interacted item (a value between 1 and 7), row duration in seconds, the sequence of items presented (0 = letter target / dog target, 1 = monkey target, 2–7 = nontarget), and the responses made by the participant (1 = selected, 0 = not selected). Action files contain timestamps for each action made by the participant, for example when selecting an item or selecting the green arrow. Summary files contain session metadata information, total hits, misses, false alarms and correct rejections, as well as Concentration Performance and Total Performance measures. Note that the structure of log files may change in the future to better reflect the needs of users.

UCancellation can be taken offline, in which case the resulting data are pushed to the server once an internet connection is established. No personally identifiable information is stored on the server and only authorized users can gain access to study-specific folders, thereby ensuring data security. Researchers who wish to test UCancellation can download Recollect the Study on the App Store (https://apps.apple.com/us/app/recollect-the-study/id1217528682) or Google Play Store (https://play.google.com/store/apps/

---

[1] The source code has since been changed to generate exactly 40 targets for every set of 10 rows.

[2] HIPAA stands for Health Insurance Portability and Accountability Act.

details?id=com.ucr.recollectstudy&hl=en_US&gl=US).
To activate UCancellation: (1) open Recollect the Study
and select NEW USER, (2) enter *let* or *pic* for Letters
or Pictures, respectively, and (3) create an anonymous
username. Note that these codes are for testing purposes
only. For more information about UCancellation setup and
data access, interested users should contact bgcresearch@
ucr.edu.

UCancellation Letters Letter stimuli resemble those used in
d2 (Brickenkamp & Zillmer, 1998), namely characters "d"
and "p" with one to four marks, arranged either individually
or in pairs above or below the letters (see Fig. 1). The par-
ticipant must scan the stimuli in a row from left to right and
select all "d"s with two marks (target) and ignore "d"s with
one, three, or four marks and all "p"s. The test involves pre-
senting at least 30 rows with 120 targets, and bonus rows are
added as long as the global time limit has not been exceeded.

UCancellation Pictures Pictures of dogs and monkeys are
presented instead of letters, some of which have inverted
colors and are reflected over the vertical axis or are pre-
sented upside down (Fig. 1). There are two types of targets:
an upright dog (tail on the left) and an upside-down monkey
(tail on right). Each target type is practiced separately in
two single blocks prior to proceeding to the mixed block.
The single blocks have a global time limit of 1 minute and
10 seconds whereas the mixed block has a global time limit
of 3 minutes and 30 seconds (thereby matching UCancella-
tion Letters). Each single block consists of a minimum of
10 rows with 40 targets whereas the mixed block contains a
minimum of 30 rows with 120 targets. Bonus rows are added
in all three block types if there is time left.

d2

Participants completed the standardized paper-and-pencil
version of d2 (Brickenkamp & Zillmer, 1998). The test
consists of one practice line and 14 test lines, each of which
consists of 47 characters "d" and "p" with 1–4 marks. The
participant must scan the lines to cross out all letters "d"
with two marks while ignoring all other distractors. Stand-
ard administration and scoring procedures were employed;
the experimenter timed each line, and after 20 seconds the
participant was asked to move to the next line. Unlike UCan-
cellation, in the d2, identical sequences are repeated every
three rows. The dependent variables used here are Concen-
tration Performance (CP), which is calculated as the number
of correctly crossed-out targets minus false alarms ($\sum$Hits
$- \sum$False alarms) and Total Performance (TP), which is
derived from the total number of items presented minus
error scores: $\sum$N $- \sum$(Misses + False alarms).

Countermanding

Countermanding is a tablet-based measure of processing
speed and executive functioning that was adapted from
Davidson et al. (2006) but using new stimuli as illustrated in
Ramani et al. (2020). The participant is instructed to tap on
one of two green buttons in response to a visual stimulus—a
picture of a dog or a monkey, displayed one at a time. For
dogs, the participant must tap on the button that is on the
same side of the screen (congruent trial); however, when the
participant sees a monkey, they must tap on the button that
is on the opposite side (incongruent trial), and thus, the par-
ticipant must inhibit a prepotent response to respond on the
same side as the visual stimulus. The visual stimuli appear
on the left or on the right randomly with equal frequency and
the interstimulus interval is 1 second (gray screen). There
are three blocks of trials, each preceded by a tutorial. An
initial block of 12 congruent trials is followed by a block
of 12 incongruent trials, both of which serve as practice,
after which a mixed block of 48 trials is presented, where
congruent and incongruent trials are randomly intermixed.
The stimuli are presented on the screen until the partici-
pant responds, with an upper time limit of 15 seconds. The
dependent measures were mean reaction times for correct
responses in the mixed block for (1) congruent trials (index
of Processing Speed), (2) incongruent trials (index of Inhibi-
tory Control), and (3) switch trials (index of Shifting), i.e.
trials for which the rule changed from congruent to incon-
gruent or vice versa (Ramani et al., 2020). Note that we do
not report or analyze reaction time (RT) difference measures
due to their generally low reliability, and because they do
not unambiguously isolate the process of interest (Miller
& Ulrich, 2013; Paap & Sawi, 2016). Since the dog and
monkey pictures used here were also used in UCancellation,
we controlled for potential performance effects with coun-
terbalancing: half of the participants did UCancellation first
and the other half did Countermanding first.

NIH EXAMINER

EXAMINER is a neuropsychological test battery that aims
to reliably and validly assess domains of executive function
(Kramer et al., 2014). We selected four tasks from form A
that enabled us to calculate a Cognitive Control (CC) com-
posite score via the EXAMINER scoring program: Flanker,
Set Shifting, Continuous Performance Task, and Antisac-
cades. All tasks were completed on a Windows desktop com-
puter. In *Flanker*, participants were instructed to focus on a
small cross in the center of the screen. After a short variable
duration, a row of five arrows was presented in the center
of the screen either above or below the fixation point. The
participant was asked to indicate whether the centrally pre-
sented arrow is pointing either to the left or right by pressing

the left or right arrow key as quickly and as accurately as possible. In congruent trials, the non-target arrows pointed in the same direction as the target arrow and in the incongruent trials they pointed in the opposite direction. The stimuli were presented in a random order with each condition presented 24 times resulting in 48 total trials. In *Set Shifting*, participants were required to match a stimulus on the top of the screen to one of two stimuli in the lower corners of the screen. In homogeneous blocks, participants performed either Task A (e.g., classifying shapes) or Task B (e.g., classifying colors). In heterogeneous blocks, participants alternated between the two tasks pseudo-randomly. There were 20 trials in each homogeneous block and 64 trials in the heterogeneous block. In the *Continuous Performance Task,* participants saw different images in the center of the screen and were asked to press the left arrow key for a target image (a white five-pointed star), responding as quickly and accurately as possible. The task consisted of 20 practice trials and 100 experimental trials, 80% of which were the target image. In *Antisaccades*, participants looked at a fixation point in the center of a computer screen and made eye movements upon seeing a laterally presented stimulus. For practice, participants completed 10 prosaccade trials in which they were instructed to look toward the stimulus. During the test, two antisaccade blocks were presented (20 trials each) in which they were asked to look in the opposite direction of the presented stimulus. An experimenter logged the direction in which the participant looked at for each trial.

We used the CC composite score as a dependent variable, which was obtained using an R script provided by the EXAMINER battery and consisted of the total Flanker score (sum of Flanker accuracy and reaction time scores), total Set Shifting score (sum of Set Shifting accuracy and reaction time scores) and Antisaccade total score (total number of correct antisaccade trials). The Continuous Performance Task was not included in the composite because this study did not include a different set of EXAMINER tasks that are needed to calculate an optional composite error score. For more details about the tasks and the outcome measures, see the EXAMINER User Manual (Kramer, 2010).

Surveys

Participants completed the following set of surveys in Qualtrics (Qualtrics, Provo, UT): (1) Demographic survey, (2) Mini-markers (Saucier, 1994), a 40-item subset of the Big Five markers of personality, (3) BIS/BAS Scales (Carver & White, 1994), a 24-item self-report survey that assesses individual differences in two motivational systems—the behavioral inhibition system (BIS) and the behavioral activation system (BAS), (4) Cognitive Failures Questionnaire (CFQ; Broadbent et al., 1982), which assesses everyday attention,

memory, and motor failures, and (4) Video Game Playing Questionnaire Version November 2019 (Green et al., 2017).

Data analysis

For purposes of comparison, we decided to use the same metrics to analyze UCancellation performance as those used for the d2 task (Brickenkamp & Zillmer, 1998), namely Total Performance and Concentration Performance. We ran bivariate correlations in IBM SPSS 24 to determine UCancellation parallel-test reliability and to examine relationships between UCancellation tasks, Countermanding, d2, and a Cognitive Control composite. Internal consistency was measured with Cronbach's alpha while split-half reliability was measured with Spearman-Brown. Outliers were removed on individual tasks if $|z| > 3$ (see below). In the UCancellation Pictures group, one outlier was removed based on Concentration Performance, hence, the final sample size was 53; no outliers were removed from the UCancellation Letters group (final sample size = 50). On the d2, three participants had missing data and one outlier was removed. For the Countermanding task, data were missing or corrupted for seven participants and two outliers were removed. For the Cognitive Control composite, data were missing for five participants and one outlier was removed. Throughout the paper, * denotes $p < 0.05$ and ** marks $p < 0.01$.

Results

Descriptive statistics

UCancellation Letters

In the first session, the number of presented rows ranged from 30 to 47 ($M = 38.90$, $SD = 4.65$), and 96% of the sample were presented with bonus rows. In the second session, the number of presented rows ranged from 32 to 52 ($M = 42.66$, $SD = 4.38$), indicating all participants were presented with bonus rows. A higher number of presented rows indicates faster processing speed, but this needs to be corrected for errors, hence we calculated Total Performance, $TP = \sum N - \sum(\text{Misses} + \text{False alarms})$, where N is the number of rows processed multiplied by 8 (number of items per row). In session 1, TP ranged from 206 to 371 ($M = 302.22$, $SD = 41.15$) while in session 2 it ranged from 252 to 413 ($M = 337.58$, $SD = 35.37$). Total Performance correlated highly with the other main outcome measure, Concentration Performance ($\sum \text{Hits} - \sum \text{False alarms}$), both in session 1 ($r = .96**$) and session 2 ($r = .97**$), which corresponds to the correlation between these two scores reported for d2 (mean $r = .93$) (Brickenkamp &

Table 2 Descriptive statistics for Concentration Performance in the first 30 rows and across all presented rows, presented separately for Letters and Pictures.

	Session 1		Session 2	
	CP: first 30 rows	CP: all rows	CP: first 30 rows	CP: all rows
Letters				
Mean (std. error)	109.98 (1.65)	143.22 (3.87)	116.58	165.20 (2.56)
95 CI of the mean	106.66–113.30	135.44–151.00	115.53–117.63	160.06–170.34
Median	113.00	144.50	116.50	163.50
Variance	136.51	748.46	13.76	326.78
Std. deviation	11.68	27.36	3.71	18.08
Minimum	77	77	108	121
Maximum	127	191	124	202
Skewness	−1.10	−0.45	−0.02	−0.09
Kurtosis	0.68	−0.35	−0.56	0.20
N	50	50	50	50
Pictures				
Mean (std. error)	110.51 (1.02)	143.04 (2.85)	115.42 (0.77)	164.53 (2.97)
95 CI of the mean	108.47–112.55	137.32–148.75	113.87–116.96	158.57–170.49
Median	111	141	115	160
Variance	54.68	429.56	31.59	467.06
Std. deviation	7.39	20.73	5.62	21.61
Minimum	89	104	101	124
Maximum	128	191	128	218
Skewness	−0.38	0.27	−0.19	0.29
Kurtosis	0.57	−0.49	−0.22	−0.57
N	53	53	53	53

Note. CP = ∑Hits − ∑False alarms. Variance: the sum of the squared distances of data value from the mean by N−1. Skewness: a measure of asymmetry of the distribution. Kurtosis: positive values indicate that the data show more extreme values than a normal distribution, whereas negative values indicate the data show less extreme outliers than a normal distribution

Zillmer, 1998). The upper part of Table 2 shows descriptive statistics for Concentration Performance in the first 30 rows and for all rows (30 rows + bonus rows). Shapiro-Wilk tests and inspection of quantile-quantile (Q-Q) plots indicated that CP in the first 30 rows in session 1 deviated from a normal distribution ($W(50) = .90*$), whereas the other three measures did not. Parallel-test reliability for CP in the first 30 rows was poor and nonsignificant ($rho = .21$), confirming that taking the whole test into account (CP: all rows) provides a more reliable measure of performance ($r = .75**$) as compared to CP in the first 30 rows. Moreover, CP in the first 30 rows of session 1 was highly negatively skewed ($< −1$), whereas CP in all rows was only moderately negatively skewed and not skewed at all in the second session. Note that sample skewness may not necessarily apply to the population. The distribution was approximately symmetric in the second session. CP for all rows in the second session was significantly higher than in the first session ($t(49) = −7.11**$).

UCancellation Pictures

In the first session, the number of presented rows ranged from 32 to 50 ($M = 38.62, SD = 4.40$), whereas in the second session it ranged from 32 to 56 ($M = 42.75, SD = 5.24$); thus in both cases all participants saw at least two bonus rows. In session 1, Total Performance ranged from 248 to 393 ($M = 301.43, SD = 35.31$) and in session 2 it ranged from 255 to 444 ($M = 336.70, SD = 41.67$). Total Performance correlated strongly with CP in session 1 ($r = .94**$) and session 2 ($r = .96**$). The bottom part of Table 2 shows descriptive statistics for Concentration Performance in the first 30 rows and in all rows (including bonus rows). Shapiro-Wilk tests indicated that the only measure that deviated from a normal distribution was CP for the first 30 rows in session 1, $W(50) = .82, p < 0.05$. Like in the Letters version, the parallel-test reliability of this measure was poor and nonsignificant ($rho = .26$) whereas CP for all rows provided a more reliable measure of performance ($rho = .73**$). CP was moderately negatively skewed in the first 30 rows of session 1 whereas it
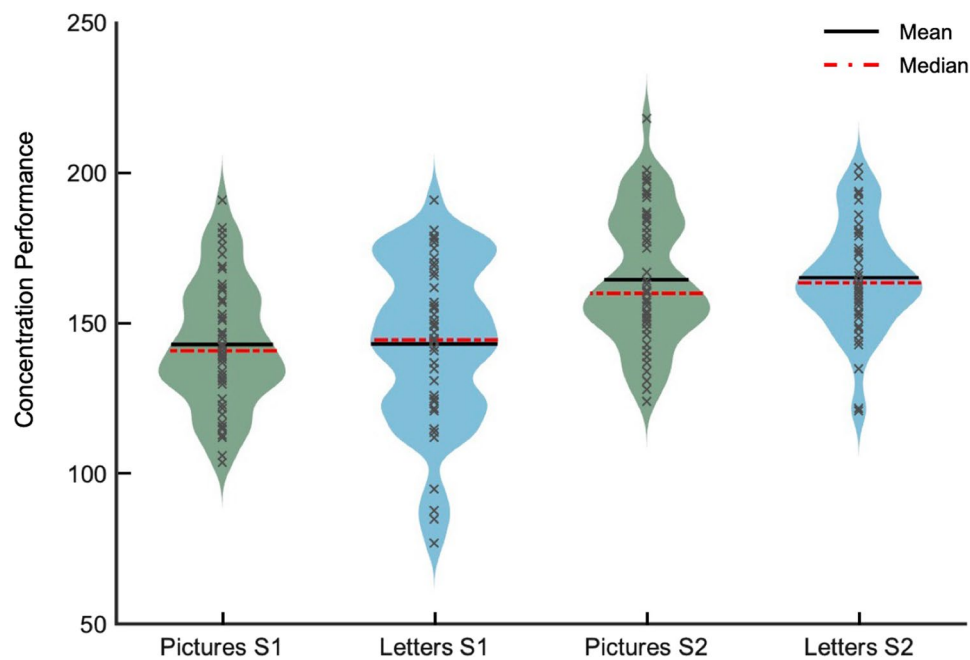
Fig. 2 Violin plot (Hoffmann 2021) of Concentration Performance in UCancellation Letters and UCancellation Pictures groups in session 1 (S1) and session 2 (S2). Wider parts of the violin plot indicate a higher probability for a data point to occur in a certain section

was approximately symmetrical when accounting for all rows (both sessions). When comparing the two sessions, CP for all rows showed significant practice effects, $t(52) = -10.32$**.

Due to poor parallel-test reliability of CP in the first 30 rows, we used CP for all rows as the dependent measure in subsequent analyses (referred to as simply "CP" hereafter).

Comparing UCancellation Letters and Pictures

Independent samples t-tests showed that there was no significant difference in Concentration Performance on Letters and Pictures in session 1 ($t(101) = 0.04$, $p = 0.97$) or in session 2 ($t(101) = 0.17$, $p = 0.87$). Likewise, there was no significant difference in Total Performance (TP) between the two tasks in session 1 ($t(101) = 0.10$, $p = 0.92$) or in session 2 ($t(101) = 0.12$, $p = 0.91$), suggesting that in this sample, Letters and Pictures yield comparable results within the same session. See Fig. 2 for a visualization of distributions and probability density of Concentration Performance for Letters and Pictures in each session.

Other tasks

Descriptive statistics for d2 Concentration Performance and Total Performance, Countermanding performance[3], and the

---

[3] Countermanding accuracy was 0.97% and 0.98% for Letters and Pictures groups, respectively.

EXAMINER Cognitive Control composite are presented in Table 3.

Reliability comparison

UCancellation Pictures showed good parallel-test reliability (Fig. 3) both in terms of CP ($r = 0.74$**, $r^2 = 0.55$) and TP ($r = 0.83$**, $r^2 = 0.69$; Letters: $r = 0.68$**, $r^2 = 0.47$). Based on inspection of scatterplots and interquartile range, two low outliers were identified and removed in the second session of Letters (data points outside of 1st quartile − 1.5*interquartile range). The resulting parallel-test reliability of Letters CP ($r = 0.71$**, $r^2 = 0.50$; $N = 48$) and Letters TP ($r = 0.76$**, $r^2 = 0.58$; $N = 48$) was comparable to that of Pictures reported above (see Fig. 3). d2 was only performed once in this study; however, TP parallel-test correlation for Pictures ($r = 0.83$) and Letters ($r = 0.76$) are comparable to the 6-hour retest reliability reported for d2 Total Performance ($r = 0.82$) in Brickenkamp and Zillmer (1998). One advantage of the UCancellation task is that a different sequence of items is generated each time, whereas the d2 has no alternate versions and even repeats the same rows of stimuli within the same test.

Parallel-test reliability was also examined for error scores and since these measures are not normally distributed, we report Spearman's rho correlations for all variables in Table 4. Reliability coefficients are shown in diagonals and

Table 3 Descriptive statistics for d2 Concentration Performance (CP), d2 Total Performance (TP), and Countermanding performance

	d2 CP	d2 TP	Countermanding Processing Speed (ms)	Countermanding Inhibitory Control (ms)	Countermanding Shifting (ms)	Cognitive Control Composite
Mean (std. error)	201.66 (4.49)	494.96 (8.10)	804.28 ms (11.57)	822.02 (11.64)	823.12 (11.07)	1.20 (0.05)
95 CI of the mean	192.76–210.56	478.89–511.03	781.31–827.24	798.91–845.13	801.13–845.10	1.10–1.31
Median	200.50	499	783.93	798.71	806.06	1.22
Variance	2012.93	6563.05	12710.52	12867.46	11645.46	0.28
Std. deviation	44.87	81.01	112.74	113.43	107.91	0.53
Minimum	73	315	617.55	653.36	638.97	0.12
Maximum	286	637	1171.58	1204.94	1152.25	2.71
Skewness	−0.17	−0.23	0.72	0.92	0.66	0.12
Kurtosis	−0.29	−0.68	0.35	0.99	0.09	−0.23
N	100	100	95	95	95	98
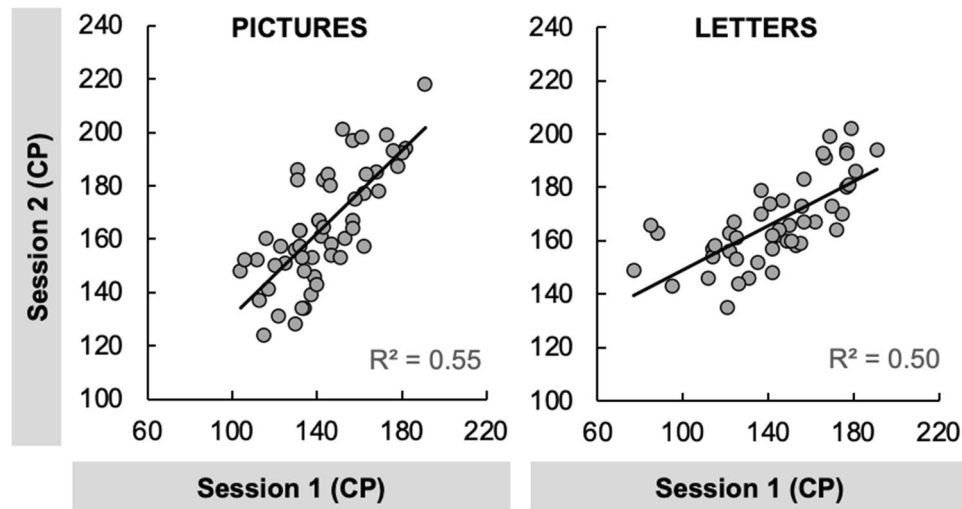
Fig. 3 Scatter plot of Concentration Performance in UCancellation Pictures ($N = 53$) and UCancellation Letters ($N = 48$) across two sessions. Two outliers were removed from the UCancellation Letters group based on performance in the second session

for the interested reader, we add intercorrelations of performance measures within session 1 (below the diagonal) and session 2 (above the diagonal). While speed-related measures showed good parallel-test reliability ($rho = .79$–$.82$**), the two types of errors, misses and false alarms, were not reliable at all ($rho = .19$–$.48$). Of the two error scores, false alarms were slightly more reliable ($rho = .40$–$.48$**), although not sufficiently so. Of note, low reliability of error scores has also been reported for the d2 Test of Attention despite the fact that the same test versions were used in the analysis (Brickenkamp & Zillmer, 1998; Steinborn et al., 2018).

Internal consistency

While the internal consistency of CP in the first 30 rows in session 1 was lower for UCancellation Pictures compared to UCancellation Letters in the first 30 rows (Pictures: $\alpha_{30} = 0.44$, $N = 53$; Letters: $\alpha_{30} = 0.80$, $N = 50$), this estimate may not be accurate because each row contains 3–5 targets, and the number of targets has a strong effect on Concentration Performance per row, which in turn affects internal consistency. This was accounted for by calculating Cronbach's alpha separately for rows that contained three, four, or five targets across five random trials within the first 30 trials of each task. As can be seen in Table 5, alpha values per row type are high and comparable across the two tasks. The highest internal consistency was observed for d2 CP ($\alpha_{14} = 0.98$, $N = 85$), which is in line with the values reported in the literature (Brickenkamp & Zillmer, 1998) and reflects the fact that the first three rows of d2 are repeated multiple times.

Table 4 Spearman's rho reliability coefficients in diagonals and intercorrelations between performance measures presented separately for Letters and Pictures

Session 1	Session 2				
	Speed	TP	CP	Misses	False alarms
Letters					
Speed	**.80****	.99**	.94**	−.23	.03
TP	.97**	**.80****	.96**	−.29*	−.04
CP	.91**	.97**	**.75****	−.36*	−.13
Misses	−.22	−.29*	−.40**	**.38***	.11
False alarms	−.08	−.22	−.30**	.42**	**.40****
Pictures					
Speed	**.79****	.99**	.92**	.19	.03
TP	.99**	**.82****	.95**	.11	−.07
CP	.91**	.93**	**.73****	.04	−.17
Misses	−.11	−.21	−.35**	**.19**	.25
False alarms	.18	.08	−.03	.34*	**.48****

Note. Speed = number of items processed; CP = hits corrected for false alarms in all rows; TP = number of items processed corrected for errors; Misses = total errors of omission; False alarms = total errors of commission. Parallel-test reliability is shown in the diagonals (in bold text). Correlation coefficients for the first testing session are shown below and for the second testing session above the diagonals. * = $p < 0.05$; ** = $p < 0.01$

Table 5 Cronbach's alpha of Concentration Performance per row type, calculated across five trials within the first 30 trials

	Alpha
Letters	
3 targets per row	0.89
4 targets per row	0.89
5 targets per row	0.91
Pictures	
3 targets per row	0.73
4 targets per row	0.84
5 targets per row	0.83

Convergent validity

Convergent validity was established for both UCancellation tasks, with the Pictures version showing somewhat higher correlations with d2 compared to Letters. Both tasks showed a significant correlation with d2 in terms of CP (Pictures: $r = 0.59$**; Letters: $r = 0.43$**) (Fig. 4) and Total Performance (Pictures: $r = 0.52$**; Letters: $r = 0.44$**); see Tables 6 and 7. Performance in the second session of both tasks correlated strongly with d2 in terms of Concentration Performance (Pictures: $r = 0.64$**; Letters: $r = 0.50$**) and Total Performance (Pictures: $r = 0.61$**; Letters: $r = 0.52$**).

In addition, both versions of the UCancellation task (in the first session) correlated significantly with NIH EXAMINER Cognitive Control in terms of Concentration Performance (Pictures: $r = 0.41$*; Letters: $r = 0.34$*) (Fig. 4) and Total Performance (Pictures: $r = 0.53$*; Letters: $r = 0.37$*), providing further evidence of the convergent validity of UCancellation. d2 also correlated significantly with Cognitive Control for Concentration Performance ($r = 0.38$*) and for Total Performance ($r = 0.34$*). These results suggest that UCancellation tasks capture a similar amount of variance related to cognitive control processes compared to the d2. In order to test this, a linear regression was conducted to predict the contributions of the UCancellation task (collapsed across Letters and Pictures versions in session 1) and d2 to Cognitive Control ($N = 95$, $R^2 = .18$). Results showed that Cognitive Control was predicted both by UCancellation CP ($B = 0.005$, $SE = 0.002$, $\beta = .228$, $p = 0.039$) and d2 CP ($B = 0.003$, $SE = 0.001$, $\beta = .266$, $p = 0.017$).

Only Concentration Performance in the Pictures version showed a significant correlation with Processing Speed as measured by the Countermanding task (Letters: $r = -0.28$; Pictures: $r = -0.46$**) and a similar pattern was observed for Total Performance (Letters: $r = -0.29$; Pictures: $r = -0.58$**). Both versions of UCancellation (and both measures, CP and TP) correlated significantly with Inhibitory Control and Shifting as measured by the Countermanding task; specifically, faster average reaction time on incongruent and switch trials was related to better performance on the UCancellation tasks (see Tables 6 and 7 and Fig. 4). A similar pattern of correlations was observed between d2 and the Countermanding measures, providing further evidence of convergent validity.

Self-report measures of personality, BIS/BAS, cognitive failures, or estimated number of hours of playing video games in the past year did not correlate with performance on UCancellation or d2 (see Table S1 in Supplementary Materials).

Discussion

Here we present data regarding the psychometric properties of a new cancellation test called UCancellation. The test runs on iOS and Android devices, can be self-administered in 7 minutes, and the resulting data are automatically scored. The participant is tasked with the goal of selecting specific targets (letters or pictures) among a row of distractors, and to do this as quickly and accurately as possible. UCancellation will not show ceiling effects as is often the case with similar speeded tasks, because new rows are presented as long as an overall time limit has not expired. Results obtained from a sample of US students suggest that UCancellation is an internally consistent and valid test of selective attention and

concentration that can be used as an alternative to the d2 Test of Attention (Brickenkamp & Zillmer, 1998).

Parallel-test reliability was assessed in two sessions conducted one week apart, at the same time of day. This procedure helped minimize the effects of sleep–wake homeostatic and internal circadian time-dependent effects on cognitive performance (Blatter & Cajochen, 2007). UCancellation Pictures showed marginally higher parallel-test reliability than UCancellation Letters for speed-related measures, and both were within a good range (*rho* above .7), albeit lower than in traditional paper-and-pencil cancellation tasks (Brickenkamp & Zillmer, 1998; Steinborn et al., 2018); however, we did not assess reliability of the paper d2 in the same population as used herein. Since the UCancellation app automatically presents new sequences of items every time the task is performed, it is important to note that here we report parallel-test reliability, which tends to be lower than test-retest reliability. While the lack of "true" test-retest reliability data presents a limitation, it is also a key strength of the current approach where participants can't learn the stimulus patterns (either explicitly or implicitly), as is possible with the d2, for which the exact same sequence of stimuli is repeated every three lines. There are several other possible reasons why UCancellation shows lower reliability than the d2. For example, it has been observed that performance on speeded tests tends to improve when an experimenter is present (Bond & Titus, 1983; Steinborn & Huestegge, 2020), which was the case for d2 in our study, and is often true for paper-and-pencil tasks but not computerized tasks. Task mode may further affect discrepancies in reliability between the two types of tasks. For example, in paper-and-pencil d2, one must scan and check 47 items per row, whereas in UCancellation and other computerized cancellation tasks, the participant is presented with fewer items per row; and may in fact elicit choice reaction time (RT) processes rather than scanning and checking. UCancellation could thus be seen as a choice-RT task with eight stimuli and two responses (tap or skip), although usually stimuli are presented sequentially in these types of tasks and not at once, as in our task.

Error scores, however, were not reliable, possibly because misses and false alarm events are rare in the tested population, and/or due to differences in how nontargets were distributed across sequences in the two sessions. This finding is in line with previous research indicating lower reliability of error scores compared to speed scores (number of items worked through) in cancellation tasks (Brickenkamp & Zillmer, 1998; Hagemeister, 2007; Steinborn et al., 2018) and supports the recommendation not to rely on error rate to index diligence. Of note, error-corrected speed scores maintain similar psychometric quality to pure speed scores (i.e. both in our data and in that reported for the d2; Steinborn et al., 2018); therefore the two measures can be used interchangeably.

Most of the change in performance between the two sessions appears to stem from practice effects. This presents a limitation in that the task is not well suited towards consecutive testing. Strong practice effects have also been reported for d2 (Bühner et al., 2006) and, more generally, in tasks requiring item-solving processes in addition to perceptual and motor processes (Blotenberg & Schmidt-Atzert, 2019). Since these effects seem to be unavoidable in speeded perceptual-motor tasks, one can apply various statistical corrections to account for them (Bruggemans et al., 1997; Maassen et al., 2009). While we have tried various tutorial structures to reduce these practice effects, this type of motor skill learning is known to be sleep dependent (Walker et al., 2002). Further work should investigate whether there is a retest time frame in which practice effects do not play a meaningful role anymore and whether test-retest effects change as a function of test length using cumulative reliability function analysis (Steinborn et al., 2018).

Due to the variable number of rows that each individual performs, as well as the changing number of targets, internal consistency was estimated for a subsample of trials that consisted of three, four, or five targets. Cronbach's alpha values were good (above 0.7) for both versions of the task. Slightly lower alpha values were obtained for Pictures compared to Letters, which might reflect the use of two distinct targets (dog, monkey) each with its own set of distractors, instead of one (letter d), which introduces more variability in performance. Internal consistency was also calculated for d2 Concentration Performance, which was very high and in line with the values reported in the literature (Brickenkamp & Zillmer, 1998). The high internal consistency of d2 is not surprising given that only the first three out of fourteen rows of d2 are unique and are thereafter repeated in the same order: the first two rows are repeated five times and the third row is repeated four times.

Convergent validity with d2 Test of Attention was established for both UCancellation tasks, with the Pictures version showing a somewhat higher correlation with d2 ($r = 0.59$) than Letters did ($r = 0.43$); however, this may in part reflect group differences. Compared to d2, UCancellation offers typical advantages of computerized tests such as standardization in administration and scoring, ease of participant recruitment if conducted remotely, reduced testing time, and the ability to obtain rich, trial-by-trial data, which are instantly available to the experimenter. That said, it should be acknowledged that any app-based task will differ substantially from paper-and-pencil variants in that it provides a substantially different visual and motor experience. While a phone or a tablet can be laid flat on a table as a paper would be, users may decide to hold it numerous different ways unbeknownst to the experimenter (if administered remotely). Performance on UCancellation will depend on how agile a person is with the use of touchscreens, which
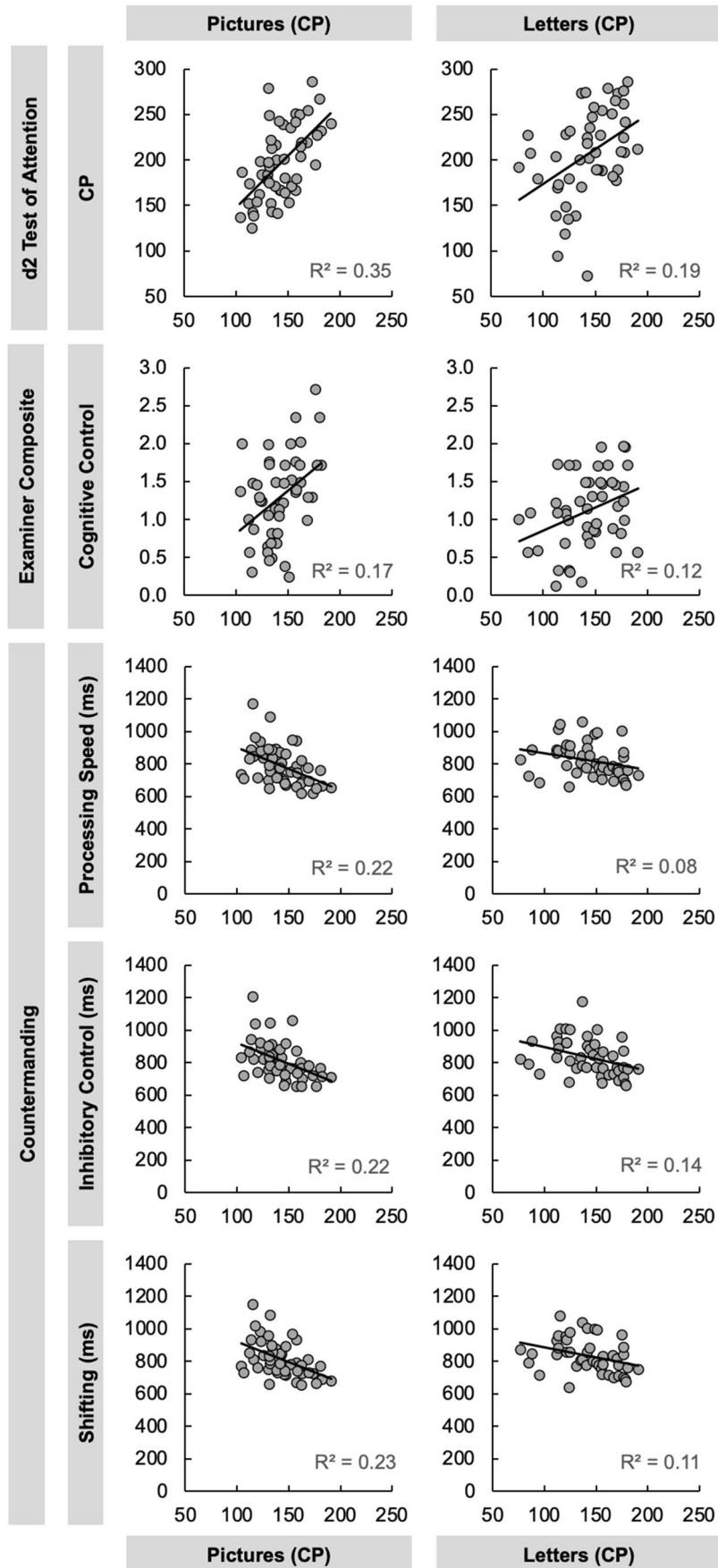
◄**Fig. 4** Scatter plots of Concentration Performance in UCancellation Pictures (left column) and UCancellation Letters (right column) in relation to d2 (row 1), the EXAMINER Cognitive Control composite (row 2) and Countermanding RT measures (rows 3–5)

may in turn affect the psychometric quality of the data (Noyes & Garland, 2008; Parks et al., 2001). Computerized tasks are also subject to freezing and crashing; however, the benefits of remote testing are clear and particularly important during times in which in-person research is not possible.

Moreover, both UCancellation tasks and d2 predicted performance on a NIH EXAMINER Cognitive Control composite to a similar extent. Likewise, both versions of UCancellation and d2 correlated significantly with Inhibitory Control and Shifting as measured by the Countermanding task, suggesting that faster performance on incongruent and switch trials is related to better performance on UCancellation and d2 tasks. Notably, Countermanding Shifting tended to show stronger correlations with performance on the Pictures task compared to the Letters task, which is in line with our prediction that the Pictures task places greater demands on switching between the two target rules. Interestingly, only the Pictures version of UCancellation, but not Letters, and d2 showed significant correlations with Processing Speed as measured by the Countermanding task (Tables 6 and 7). The results should be interpreted with caution given that RT-based correlations depend on task-specific characteristics as well as population characteristics, for example general and task-specific processing time variability, and their correlation (Miller & Ulrich, 2013). These preliminary findings need to be explored further in a larger sample size and if correlations are made that involve RT as a dependent measure, individual differences in reaction time need to be accounted for.

There was no evidence to suggest that self-report measures such as personality attributes, BIS/BAS scales, prevalence of cognitive failures, or estimated number of hours of playing video games in the past year are associated with performance on UCancellation, d2, or the Cognitive Control composite (see Table S2 in Supplementary Materials). The main purpose of including these surveys was to verify that any potential relations with cognitive performance are stable across tests. For example, if d2 performance were to correlate with one of the administered surveys, we would expect to see a similar correlation between UCancellation and that survey. However, none of the measures showed significant correlations with any of the surveys. The only exception was Countermanding Inhibitory Control, which correlated with self-reported number of hours spent playing video games in the past year, such that more video game playing was associated with faster reaction times on incongruent trials, which is line with previous findings (Dye et al., 2009; Hutchinson et al., 2016).

A limitation of the current dataset is gender imbalance in favor of women because we relied on a convenience sample consisting of mostly Psychology/Education students where women tend to be overrepresented. Note that in a large sample of university students, no significant differences between men and women were found in selective attention and mental concentration as measured by the d2 Test of Attention (Fernández-Castillo & Caurcel, 2015). Likewise, age, sex, and level of education did not affect cancellation performance in a large and demographically stratified group of healthy adults (Benjamins et al., 2019); therefore, we do not expect that this imbalance would affect the present results. Nevertheless, a larger, more representative sample including a wider age range and educational level would be needed as a next step in the validation of UCancellation.

As online testing becomes more and more ubiquitous, it is becoming increasingly important to develop psychometrically validated app-based tools that can be delivered on participants' own devices and that provide researchers, clinicians, and other professionals with accurate and secure data. Recent evidence suggests that tablet-based measures of cognitive abilities can provide a quick and reliable way of collecting large datasets (Bignardi et al., 2021). As opposed to other freely available cancellation tasks, the UCancellation app is optimized for performance on different screen sizes and comes with access to custom-built software for instant access to the data, which can be stored on a HIPAA-compliant server. The test can be taken offline, in which case the data are uploaded to the server once an internet connection is established. These features ensure that remote online testing is not only possible, but also convenient. The app can be easily modified to support different stimulus sets, such as different sets of letters or pictures that are deemed appropriate for a certain population. The use of pictures instead of letters can increase culture-fairness. Pictures of animals also may be more engaging than letters, particularly for children, although this hypothesis is yet to be tested. Importantly, researchers can administer both versions of UCancellation, verbal and pictorial, in order to measure attention as a general ability (Wühr & Ansorge, 2020; Flehmig et al., 2007), or to examine content-facet factors.

We note that UCancellation may be more appropriate for diverse populations such as children or older adults because the stimuli are much larger than those used in d2 and d2-R. While earlier versions of UCancellation also displayed multiple rows on each screen, iterative development and piloting over the past 2 years revealed that this type of presentation is suboptimal for smaller screens. Instead, we decided to present one row at a time, allowing for larger items that are easily scalable across various screen dimensions. The advantage to performance is both visual and motor, as poor performance on other versions of cancellation tasks can easily be confounded by poor motor skills or low vision. On the other

Table 6 Pearson correlations between Concentration Performance (UCancellation Letters, UCancellation Pictures, and d2), the Cognitive Control (CC) Composite, and Countermanding Performance

	1	2	3	4	5	6	7
1. Letters CP	1						
2. Pictures CP	N/A	1					
3. d2 CP	0.43**	0.59**	1				
4. CC Composite	0.34*	0.41**	0.38**	1			
5. Countermanding processing speed	−0.28	−0.46**	−0.35**	−0.51**	1		
6. Countermanding inhibitory control	−0.37*	−0.47**	−0.42**	−0.55**	0.91**	1	
7. Countermanding shifting	−0.33*	−0.48**	−0.39**	−0.55**	0.95**	0.94**	1

Note. CP = ∑Hits − ∑False alarms in all UCancellation rows including bonus rows. CP = ∑Hits − ∑False alarms in all d2 rows. * = $p < 0.05$; ** = $p < 0.01$. Correlations that do not involve UCancellation performance measures were calculated using the entire sample

Table 7 Pearson correlations between Total Performance (UCancellation Letters, UCancellation Pictures, and d2), the Cognitive Control (CC) Composite, and Countermanding Performance

	1	2	3	4	5	6	7
1. Letters TP	1						
2. Pictures TP	N/A	1					
3. d2 TP	0.44**	0.52**	1				
4. CC Composite	0.37*	0.53**	0.34**	1			
5. Countermanding processing speed	−0.29	−0.58**	−0.31**	−0.51**	1		
6. Countermanding inhibitory control	−0.41**	−0.58**	−0.40**	−0.55**	0.91**	1	
7. Countermanding shifting	−0.35*	−0.60**	−0.36**	−0.55**	0.95**	0.94**	1

Note. TP = ∑N − ∑(Misses + False alarms). * = $p < 0.05$; ** = $p < 0.01$. Correlations that do not involve UCancellation performance measures were calculated using the entire sample

hand, this type of line-by-line layout (with only eight items per line) does not lend well to measuring search organization, typically measured as the distance between consecutive cancellations (Dalmaijer et al., 2015), and further, it eliminates the distraction provided by non-target rows. It also does not allow for measurement of spatial components of cancellation responses as an indicator of spatial neglect (Huygelier & Gillebert, 2020). Nevertheless, in the interest of being able to do the task on smartphones, we opted for line-by-line presentation, thereby making the app more accessible.

All in all, the results of the present study suggest that UCancellation is a valid and reliable cancellation measure that can be used to estimate selective attention and concentration performance in adults. Since it is freely available and can be used both for in-person and remote assessment, its use extends beyond basic research to inform neuropsychological assessment in clinical populations that experience executive function difficulties. Like most cancellation tasks, UCancellation taps into multiple cognitive processes such as selective attention, concentration, inhibitory control, and processing speed. To map how and to what extent these processes are engaged while performing the task (or another, related task), more data are needed that will support factor analytic and/or structural equation modelling approaches.

References

Bates, M.E. and Lemay, E.P. 2004. The d2 Test of attention: construct validity and extensions in scoring techniques. *Journal of the International Neuropsychological Society* 10(3), 392–400.

Benjamins, J.S., Dalmaijer, E.S., Ten Brink, A.F., Nijboer, T.C.W. and Van der Stigchel, S. 2019. Multi-target visual search organisation across the lifespan: cancellation task performance in a large and demographically stratified sample of healthy adults. *Neuropsychology, Development, and Cognition. Section B, Aging, Neuropsychology and Cognition* 26(5), 731–748.

Bignardi, G., Dalmaijer, E.S., Anwyl-Irvine, A. and Astle, D.E. 2021. Collecting big data with small screens: Group tests of

children's cognition with touchscreen tablets are reliable and valid. *Behavior Research Methods* 53(4), 1515–1529.

Blatter, K. and Cajochen, C. 2007. Circadian rhythms in cognitive performance: methodological constraints, protocols, theoretical underpinnings. *Physiology & Behavior* 90(2–3), 196–208.

Blotenberg, I. and Schmidt-Atzert, L. (2019). On the locus of the practice effect in sustained attention tests. *Journal of Intelligence,* 7(2), 12. https://doi.org/10.3390/jintelligence7020012

Bond, C.F. and Titus, L.J. 1983. Social facilitation: A meta-analysis of 241 studies. *Psychological Bulletin* 94(2), 265–292.

Brickenkamp, R. and Zillmer, E. 1998. Test d2: Concentration-Endurance Test. CJ Hogrefe.

Broadbent, D.E., Cooper, P.F., FitzGerald, P. and Parkes, K.R. 1982. The Cognitive Failures Questionnaire (CFQ) and its correlates. *The British Journal of Clinical Psychology* 21 (Pt 1), 1–16.

Brodeur, D.A. and Pond, M. 2001. The development of selective attention in children with attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology* 29(3), 229–239.

Bruggemans, E.F., Van de Vijver, F.J. and Huysmans, H.A. 1997. Assessment of cognitive deterioration in individual patients following cardiac surgery: correcting for measurement error and practice effects. *Journal of Clinical and Experimental Neuropsychology* 19(4), 543–559.

Bühner, M., Ziegler, M., Bohnes, B. and Lauterbach, K. 2006. Übungseffekte in den TAP Untertests Test Go/Nogo und Geteilte Aufmerksamkeit sowie dem Aufmerksamkeits-Belastungstest (d2). *Zeitschrift für Neuropsychologie* 17(3), 191–199.

Carver, C.S. and White, T.L. 1994. Behavioral inhibition, behavioral activation, and affective responses to impending reward and punishment: The BIS/BAS Scales. *Journal of Personality and Social Psychology* 67(2), 319–333.

Commodari, E. and Guarnera, M. 2008. Attention and aging. *Aging Clinical and Experimental Research* 20(6), 578–584.

Dalmaijer, E.S., Van der Stigchel, S., Nijboer, T.C.W., Cornelissen, T.H.W. and Husain, M. 2015. CancellationTools: All-in-one software for administration and analysis of cancellation tasks. *Behavior Research Methods* 47(4), 1065–1075.

Davidson, M.C., Amso, D., Anderson, L.C. and Diamond, A. 2006. Development of cognitive control and executive functions from 4 to 13 years: evidence from manipulations of memory, inhibition, and task switching. *Neuropsychologia* 44(11), 2037–2078.

Della Sala, S., Laiacona, M., Spinnler, H. and Ubezio, C. 1992. A cancellation test: its reliability in assessing attentional deficits in Alzheimer's disease. *Psychological Medicine* 22(4), 885–901.

Duchek, J.M., Balota, D.A., Tse, C.-S., Holtzman, D.M., Fagan, A.M. and Goate, A.M. 2009. The utility of intraindividual variability in selective attention tasks as an early marker for Alzheimer's disease. *Neuropsychology* 23(6), 746–758.

Dye, M.W.G., Green, C.S. and Bavelier, D. 2009. Increasing Speed of Processing With Action Video Games. *Current Directions in Psychological Science* 18(6), 321–326.

Ferber, S. and Karnath, H.O. 2001. How to assess spatial neglect--line bisection or cancellation tasks? *Journal of Clinical and Experimental Neuropsychology* 23(5), 599–607.

Fernández-Castillo, A. and Caurcel, M.J. 2015. State test-anxiety, selective attention and concentration in university students. *International Journal of Psychology : Journal International de Psychologie* 50(4), 265–271.

Flehmig, H. C., Steinborn, M., Langner, R., Scholz, A., and Westhoff, K. (2007). Assessing intraindividual variability in sustained attention: Reliability, relation to speed and accuracy, and practice effects. *Psychology Science,* 49(2), 132.

Gauthier, L., Dehaut, F., and Joanette, Y. (1989). The bells test: a quantitative and qualitative test for visual neglect. *International Journal of Clinical Neuropsychology,* 11(2), 49–54.

Green, C.S. and Bavelier, D. 2003. Action video game modifies visual selective attention. *Nature* 423(6939), 534–537.

Green, C.S., Kattner, F., Eichenbaum, A., et al. 2017. Playing some video games but not others is related to cognitive abilities: A critique of unsworth et al. (2015). *Psychological Science* 28(5), 679–682.

Hagemeister, C. 2007. How useful is the power law of practice for recognizing practice in concentration tests? *European Journal of Psychological Assessment* 23(3), 157–165.

Halligan, P. W., Marshall, J. C., and Wade, D. T. (1989). Visuospatial neglect: underlying factors and test sensitivity. *The Lancet,* 334(8668), 908–911.

Hoffmann, H. (2022). Violin Plot (https://www.mathworks.com/matlabcentral/fileexchange/45134-violin-plot), MATLAB Central File Exchange. Retrieved March 10, 2021.

Huang, H.-C. and Wang, T.-Y. 2009. Stimulus effects on cancellation task performance in children with and without dyslexia. *Behavior Research Methods* 41(2), 539–545.

Hutchinson, C.V., Barrett, D.J.K., Nitka, A. and Raynes, K. 2016. Action video game training reduces the Simon Effect. *Psychonomic Bulletin & Review* 23(2), 587–592.

Huygelier, H. and Gillebert, C.R. 2020. Quantifying egocentric spatial neglect with cancellation tasks: A theoretical validation. *Journal of Neuropsychology* 14(1), 1–19.

Jonkman, L.M. 2005. Selective Attention Deficits in Children with Attention Deficit Hyperactivity Disorder. In: Gozal, D. and Molfese, D. L. eds. *Attention deficit hyperactivity disorder*. Humana Press, 255–275.

Kramer, J. H. (2010). Executive abilities: measures and instruments for neurobehavioral evaluation and research (EXAMINER) User Manual 3.6. . Available at: https://memory.ucsf.edu/sites/memory.ucsf.edu/files/wysiwyg/EXAMINER_UserManual_3.6r.pdf. Accessed 20 October 2021.

Kramer, J.H., Mungas, D., Possin, K.L., et al. 2014. NIH EXAMINER: conceptualization and development of an executive function battery. *Journal of the International Neuropsychological Society* 20(1), 11–19.

Lee, P., Lu, W.-S., Liu, C.-H., Lin, H.-Y. and Hsieh, C.-L. 2018. Test-Retest Reliability and Minimal Detectable Change of the D2 Test of Attention in Patients with Schizophrenia. *Archives of Clinical Neuropsychology* 33(8), 1060–1068.

Levinoff, E.J., Li, K.Z.H., Murtha, S. and Chertkow, H. 2004. Selective attention impairments in Alzheimer's disease: evidence for dissociable components. *Neuropsychology* 18(3), 580–588.

Maassen, G.H., Bossema, E. and Brand, N. 2009. Reliable change and practice effects: outcomes of various indices compared. *Journal of Clinical and Experimental Neuropsychology* 31(3), 339–352.

Maddox, W.T., Filoteo, J.V., Delis, D.C. and Salmon, D.P. 1996. Visual selective attention deficits in patients with Parkinson's disease: A quantitative model-based approach. *Neuropsychology* 10(2), 197–218.

McDowd, J.M. and Filion, D.L. 1992. Aging, selective attention, and inhibitory processes: A psychophysiological approach. *Psychology and Aging* 7(1), 65–71.

Miller, J. and Ulrich, R. 2013. Mental chronometry and individual differences: modeling reliabilities and correlations of reaction time means and effect sizes. *Psychonomic Bulletin & Review* 20(5), 819–858.

Noyes, J.M. and Garland, K.J. 2008. Computer- vs. paper-based tasks: are they equivalent? *Ergonomics* 51(9), 1352–1375.

Paap, K.R. and Sawi, O. 2016. The role of test-retest reliability in measuring individual and group differences in executive functioning. *Journal of Neuroscience Methods* 274, 81–93.

Parks, S., Bartlett, A., Wickham, A. and Myors, B. 2001. Developing a computerized test of perceptual/clerical speed. *Computers in Human Behavior* 17(1), 111–124.

Prabhakaran, R., Kraemer, D.J.M. and Thompson-Schill, S.L. 2011. Approach, avoidance, and inhibition: personality traits predict cognitive control abilities. *Personality and Individual Differences* 51(4), 439–444.

Ramani, G.B., Daubert, E.N., Lin, G.C., Kamarsu, S., Wodzinski, A. and Jaeggi, S.M. 2020. Racing dragons and remembering aliens: Benefits of playing number and working memory games on kindergartners' numerical knowledge. *Developmental Science* 23(4), p. e12908.

Rivera, D., Salinas, C., Ramos-Usuga, D., et al. 2017. Concentration Endurance Test (d2): Normative data for Spanish-speaking pediatric population. *NeuroRehabilitation* 41(3), 661–671.

Robertson, I.H., Manly, T., Andrade, J., Baddeley, B.T. and Yiend, J. 1997. "Oops!": performance correlates of everyday attentional failures in traumatic brain injured and normal subjects. *Neuropsychologia* 35(6), 747–758.

Rorden, C. and Karnath, H.-O. 2010. A simple measure of neglect severity. *Neuropsychologia* 48(9), 2758–2763.

Saucier, G. 1994. Mini-markers: a brief version of Goldberg's unipolar big-five markers. *Journal of Personality Assessment* 63(3), 506–516.

Spaccavento, S., Marinelli, C.V., Nardulli, R., et al. 2019. Attention Deficits in Stroke Patients: The Role of Lesion Characteristics, Time from Stroke, and Concomitant Neuropsychological Deficits. *Behavioural neurology* 2019, 7835710.

Steinborn, M.B. and Huestegge, L. 2020. Socially alerted cognition evoked by a confederate's mere presence: analysis of reaction-time distributions and delta plots. *Psychological Research* 84(5), 1424–1439.

Steinborn, M.B., Langner, R., Flehmig, H.C. and Huestegge, L. 2018. Methodology of performance scoring in the d2 sustained-attention test: Cumulative-reliability functions and practical guidelines. *Psychological Assessment* 30(3), 339–357.

Stevens, C. and Bavelier, D. 2012. The role of selective attention on academic foundations: a cognitive neuroscience perspective. *Developmental cognitive neuroscience* 2 Suppl 1, S30–48.

Uttl, B. and Pilkenton-Taylor, C. 2001. Letter cancellation performance across the adult life span. *The Clinical Neuropsychologist* 15(4), 521–530.

van Steenbergen, H., Band, G.P.H. and Hommel, B. 2009. Reward counteracts conflict adaptation. Evidence for a role of affect in executive control. *Psychological Science* 20(12), 1473–1477.

Walker, M.P., Brakefield, T., Morgan, A., Hobson, J.A. and Stickgold, R. 2002. Practice with sleep makes perfect: sleep-dependent motor skill learning. *Neuron* 35(1), 205–211.

Wang, T.-Y., Huang, H.-C. and Huang, H.-S. 2006. Design and implementation of cancellation tasks for visual search strategies and visual attention in school children. *Computers & Education* 47(1), 1–16.

Wassenberg, R., Hendriksen, J.G.M., Hurks, P.P.M., et al. 2008. Development of inattention, impulsivity, and processing speed as measured by the d2 Test: results of a large cross-sectional study in children aged 7–13. *Child Neuropsychology* 14(3), 195–210.

Weintraub, S. and Mesulam M.-M. 1985. Mental state assessment of young and elderly adults in behavioral neurology. In: Edited by M.-M. Mesulam, Principles of behavioral neurology. Philadelphia: F. A. Davis, pp. 71–123

Wühr, P., and Ansorge, U. (2020). Do left-handers outperform right-handers in paper-and-pencil tests of attention?. *Psychological Research, 84*(8), 2262–2272.

Zanto, T.P. and Gazzaley, A. 2004. Selective attention and inhibitory control in the aging brain. In: Cabeza, R., Nyberg, L., and Park, D. C. eds. *Cognitive neuroscience of aging: Linking cognitive and cerebral aging*. Oxford University Press, 207–234.

Ziino, C. and Ponsford, J. 2006. Selective attention deficits and subjective fatigue following traumatic brain injury. *Neuropsychology* 20(3), 383–390.