Psychological Research
https://doi.org/10.1007/s00426-020-01348-7

ORIGINAL ARTICLE

Incremental validity of placekeeping as a predictor of multitasking

Alexander P. Burgoyne[1] · **David Z. Hambrick**[1] · **Erik M. Altmann**[1]

Received: 11 September 2019 / Accepted: 17 April 2020

Abstract

Multitasking is ubiquitous in everyday life, which means there is value in developing measures that predict successful multitasking performance. In a large sample ($N = 404$ contributing data), we examined the predictive and incremental validity of *placekeeping*, which is the ability to perform a sequence of operations in a certain order without omissions or repetitions. In the context of multitasking, placekeeping should play a role in the performance of procedural subtasks and the interleaving of subtasks that interrupt each other. Regression analyses revealed that placekeeping ability accounted for 11% of the variance in multitasking performance, and had incremental validity relative to each of a diverse set of cognitive abilities (working memory capacity, fluid intelligence, perceptual speed, and crystallized intelligence). The predictive validity of placekeeping for multitasking was stable across samples of performance and robust to placekeeping practice. Broader measures of performance on our placekeeping task accounted for 21% of the variance in multitasking performance and had incremental validity relative to an estimate of psychometric *g*. The results provide evidence that placekeeping is a distinct cognitive ability with its own specific role to play in multitasking, and raise the possibility that measures of placekeeping ability could have utility in selecting personnel for occupations that require certain kinds of multitasking, such as interleaving of procedures.

Introduction

Multitasking is ubiquitous in everyday life. As a case in point, a current occupational database lists over 900 professions that require "the ability to shift back and forth between two or more activities or sources of information" (National Center for O*Net Development, 2019). When performance is under conditions of distraction or interference, the task performer must suspend one task to perform another before returning to the first. An example is preparing a complicated meal; one must attend to multiple dishes so as not to overcook them and remember which have already been seasoned. Although multitasking failures may be merely frustrating in a cooking task, in other settings they can be costly. For example, a medical practitioner who suspends treatment of one patient to attend to a second patient must remember where they left off with the first; otherwise they might skip a step in the procedure or administer the same medicine twice. Thus, identifying which cognitive abilities predict successful multitasking performance has important implications for personnel selection across a wide range of occupations.

Here, we focus on the relationship between multitasking and a construct we call *placekeeping* (Altmann, Trafton, & Hambrick, 2014). We define placekeeping as the ability to perform a sequence of steps in a specified order without skipping or repeating steps. Placekeeping plays a role in any task with constraints on the order of operations that the task performer must adhere to, and thus should support a wide range of higher-order cognitive activities. An obvious example is performance of procedures, which consist of steps to be performed in a specified order. A less obvious example is event counting, which involves keeping one's place in the sequence of integers while contending with distracting factors such as variability in event timing (Carlson & Cassenti, 2004). A perhaps even less obvious example is problem solving, where placekeeping supports exploration of all candidate solutions (i.e., without omissions) without unproductive exploration of failed ones (i.e., without repetitions; Hambrick, Burgoyne, & Altmann, 2020). Problem solving is, in turn, a basis of fluid intelligence, and placekeeping is in fact correlated with fluid intelligence (Hambrick & Altmann, 2015; Hambrick, Altmann, & Burgoyne, 2018), even more

✉ Alexander P. Burgoyne
burgoyn4@gmail.com

[1] Department of Psychology, Michigan State University, East Lansing, MI 48824, USA

so than working memory capacity (Burgoyne, Hambrick, & Altmann, 2019).

Placekeeping should also support performance in particular kinds of multitasking environments. In an environment that requires the interleaving of different procedures, or performance of a single procedure with frequent distractions or interruptions, placekeeping plays a role when the task performer resumes a procedure on which progress has been temporarily suspended. Interruptions as short as a few seconds increase the rate at which people incorrectly resume the interrupted task (Altmann et al., 2014), which means that even a brief shift of attention to advance, or just monitor, a different task can affect placekeeping accuracy. Not all multitasking environments meet these conditions, but many routine and occupational tasks do, including complicated meal preparation (or line cooking), emergency medical procedures, and any other procedure performed in an asynchronous, dynamic environment.

In the present study, we measure placekeeping ability using the UNRAVEL task (Altmann et al., 2014). In this task, participants perform a fixed number of steps in a prescribed order, continuing with the first step once they reach the last, generating continuous performance. Performance is interrupted periodically with a distractor task, and after each interruption participants must remember the step they performed before the interruption in order to resume with the correct step. UNRAVEL is not designed to simulate any particular occupational task, but instead to capture placekeeping requirements common to a wide range of occupational tasks, as well as the role of the distractions and interruptions that make placekeeping challenging in many of the same task environments that involve performance of procedures.

In previous research, we found that the UNRAVEL task is suitable for studying individual differences in placekeeping ability, with substantial variability in performance and acceptable or better reliability (coefficient alpha > .70; Hambrick & Altmann, 2015). We also found that UNRAVEL demonstrated convergent validity with respect to another measure of placekeeping ability (average $r = .34$), suggesting that there is an underlying ability to be measured, and discriminant validity with respect to measures of working memory capacity (average $r = .13$) and fluid intelligence (average $r = .18$), consistent with placekeeping being a distinct cognitive ability (Burgoyne et al., 2019).

On a theoretical level, a cognitive model we developed accounts for placekeeping in terms of an interplay of short-term and long-term memory representations (Altmann & Hambrick, 2017; Altmann & Trafton, 2015; Altmann, Trafton, & Hambrick, 2017). Short-term memory stores episodic traces of past performance, and long-term memory stores the correct sequence of steps. The system hypothetically keeps its place in the sequence by using an episodic memory of the most recently performed step to "look up" the next step in the long-term representation of the sequence. In traditional working memory tasks, by contrast, there is no obvious task-related need for long-term memory representations, helping to explain why placekeeping shows discriminant validity with respect to working memory capacity (Burgoyne et al., 2019).

The task we use to measure multitasking ability is SynWin (Elsmore, 1994). In SynWin, performers try to maximize their score by coordinating performance of a math subtask, a memory subtask, and auditory and visual monitoring subtasks. The math subtask in particular is a multistep procedure that extends in time and needs to be suspended periodically for the performer to attend to the other subtasks, which all run asynchronously. In addition, placekeeping may also play a role more globally, by helping the performer maintain a strategy for interleaving the different subtasks. SynWin participants develop strategies for interleaving the various subtasks in ways that maximize point payoffs, as revealed by manipulations of payoff parameters (Hambrick, Oswald, Darowski, Rench, & Brou, 2010; Wang, Proctor, & Pick, 2007). Any such strategy is a procedure in the sense that it imposes a sequential structure on choices about which subtask to attend or respond to next. Thus, placekeeping ability may play a role at multiple levels of SynWin performance.

In a previous study, we found no relationship between UNRAVEL and SynWin performance (Hambrick & Altmann, 2015). However, in that study we used only one measure of UNRAVEL performance as a placekeeping indicator, and administered a shortened version of SynWin that may not have been long enough for participants to develop strategies for interleaving subtasks. Our sample size was also modest. Thus, given theoretical reasons to think there is a relationship between the two tasks, we tested again for it here, using a second placekeeping measure (response time, versus just placekeeping errors), a longer SynWin administration (25 min, versus 10 min), and a substantially larger sample ($N = 404$ contributing data, versus $N = 132$).

To assess the relationship between placekeeping and multitasking, we performed two analytical steps. In the first, we estimated predictive validity—the proportion of variance in multitasking explained by placekeeping. In the second, we estimated incremental validity—the amount of variance in multitasking explained by placekeeping above and beyond the variance explained by four other cognitive abilities: working memory capacity, fluid intelligence, perceptual speed, and crystallized intelligence.

We chose these four comparison abilities because they are widely studied and diverse enough that together they represent an estimate of psychometric *g*. In terms of McGrew's (2009) characterizations, working memory capacity is the ability to maintain awareness of information in the immediate situation, and fluid intelligence reflects the use of controlled processes to solve novel problems. In contrast, perceptual speed, which is the ability to make simple decisions about stimuli, emphasizes perceptual processes rather than memory and reasoning. Finally, crystallized intelligence, which refers to breadth and depth of acquired knowledge, is the complement of fluid intelligence in the original dichotomous conception of mental ability (Cattell, 1943). Two of the abilities (working memory capacity and fluid intelligence) are of particular interest because there are reasons to think they might overlap completely with placekeeping, as we address in the Discussion.

If placekeeping has incremental validity for multitasking relative to all four of these comparison abilities—and relative to working memory capacity and fluid intelligence in particular—this would be evidence that placekeeping is a distinct process with its own role to play in multitasking. In a final analysis, summarized in the Discussion and developed in detail in the Appendix, we take a more empirical approach and examine the predictive and incremental validity for multitasking of the UNRAVEL task as a whole, including non-placekeeping measures, to assess its prospects for helping to predict multitasking aptitude.

Methods

Participants

Our initial sample consisted of 548 undergraduate students recruited through the participant pool at Michigan State University. Of these, we excluded 102 who did not return for the second testing session, 3 who did not complete UNRAVEL, and 20 whose placekeeping error rate on UNRAVEL was not significantly better than chance-level performance. Of the remaining 423, we excluded 3 who did not complete SynWin, and 16 who had outlying scores on measures other than UNRAVEL, where an outlier on a measure is defined as a score differing by more than 3.5 standard deviations from the sample mean for that measure. Data from the remaining 404 participants were submitted to analysis.

Procedure

Participants performed tasks during two sessions on separate days, each lasting approximately 1.5 h. The order of tasks was fixed to avoid participant × order interactions, as is standard in individual difference research. In the first session, the tasks were UNRAVEL, symmetry span, and operation span. In the second session, the tasks were SynWin, Letter Sets, Vocabulary, Pattern Comparison, Raven's Progressive Matrices, Reading Comprehension, Letter Comparison, Number Comparison, and Number Series. Participants were given a short break following UNRAVEL, SynWin, and Raven's Progressive Matrices. All tests were administered via computer.

Unrelated to the present research question, participants were randomly assigned to complete one of two versions of UNRAVEL that differed in the accessibility of a help function that displayed the list of choice rules (described below). In one condition participants had unlimited access to this function, whereas in the other condition access was limited to 50 uses. This experimental manipulation interacted with none of the results presented here and is not considered further.

Placekeeping

We used the UNRAVEL task (Altmann et al., 2014) to measure placekeeping performance. Participants perform a sequence of seven steps in an order defined by the word UNRAVEL. Each letter of the word mnemonically identifies one step to perform, and when the participant reaches L they start over with U, generating continuous performance.

On a given trial, the participant performs one step of the UNRAVEL sequence on a randomly generated, multidimensional stimulus. Performing a step involves making a two-alternative forced choice decision about one dimension of the stimulus. The stimulus contains no information about the correct step to perform, so participants must remember which step they are on. The participant registers their response by pressing a key on the keyboard. The response terminates the trial, and on the next trial a new stimulus is presented and the participant moves on to the next step of the sequence. Figure 1 shows two sample stimuli (Panel a), the choice rules for each of the seven steps (Panel b), and the key the participant should press according to each rule applied to the sample stimuli (Panel b).

Performance is periodically interrupted by a typing task, which onsets immediately after the response to a trial. Figure 1 shows a sample stimulus for the interrupting task (Panel c). The participant must type the "code" correctly and then press the Return key. The code consists of the 14 candidate responses for the UNRAVEL steps, presented in random order. As the participant types, the typed characters appear in the gray box to the right of the code. The participant must ultimately type the code correctly; if there is an error in what they typed when they press Return, the gray box is cleared and they must try again. Two such codes are presented in succession during each interruption. The interruption ends when the participant presses Return after the second correctly-typed code. After the interruption, a new stimulus for the primary task appears immediately and the participant tries to resume their place in the UNRAVEL sequence where he or she left off before the interruption.

An experimental session consisted of an introductory phase followed by four test blocks. During the introductory

(a) Sample stimuli for UNRAVEL task:

(b) Choice rules and candidate responses for UNRAVEL task, and responses to the stimuli in (a):

Step	Candidate responses		Choice rules	Responses to Stimulus 1	Responses to Stimulus 2
U	u	i	character is Underlined or in Italics	u	i
N	n	f	letter is Near to or Far from start of alphabet	n	f
R	r	y	character is Red or Yellow	r	y
A	a	b	character is Above or Below the box	a	b
V	v	c	letter is Vowel or Consonant	v	c
E	e	o	digit is Even or Odd	o	e
L	l	m	digit is Less than or More than 5	l	l

(c) Sample stimulus for the transcription task:

Fig. 1 UNRAVEL. **a** Two sample stimuli for the UNRAVEL task (the A is presented in red, and the *2* in yellow). **b** Response mappings for the UNRAVEL task, along with responses for the two sample stimuli shown in panel **a**. **c** Sample stimulus for the transcription task

phase the computer presented task instructions and a series of practice trials and practice interruptions. During this phase the computer required correct responses for trials as well as interruptions. Test blocks each contained exactly 10 interruptions, dividing the block into 11 runs of trials. Each run contained an average of 6 trials, with the exact number randomized to make the occurrence of an interruption unpredictable (Altmann et al., 2014). Thus, on average there were 66 trials per block. After a block, participants received feedback on their performance and had a chance to rest.

Our placekeeping measures were the placekeeping error rate and the response time on correct trials. A placekeeping error occurs when the participant performs the wrong step in the sequence, relative to the step performed on the previous trial.[1] Response time spans from stimulus onset to the response on a trial, and thus includes all cognitive operations required to perform a trial, including the placekeeping operations required to select the next step. The two measures are

[1] For example, if the participant performs the U, R, and A steps on successive trials, a placekeeping error occurs on the R trial, because the correct step would have been N. However, the A trial is correct, because A follows R in UNRAVEL.

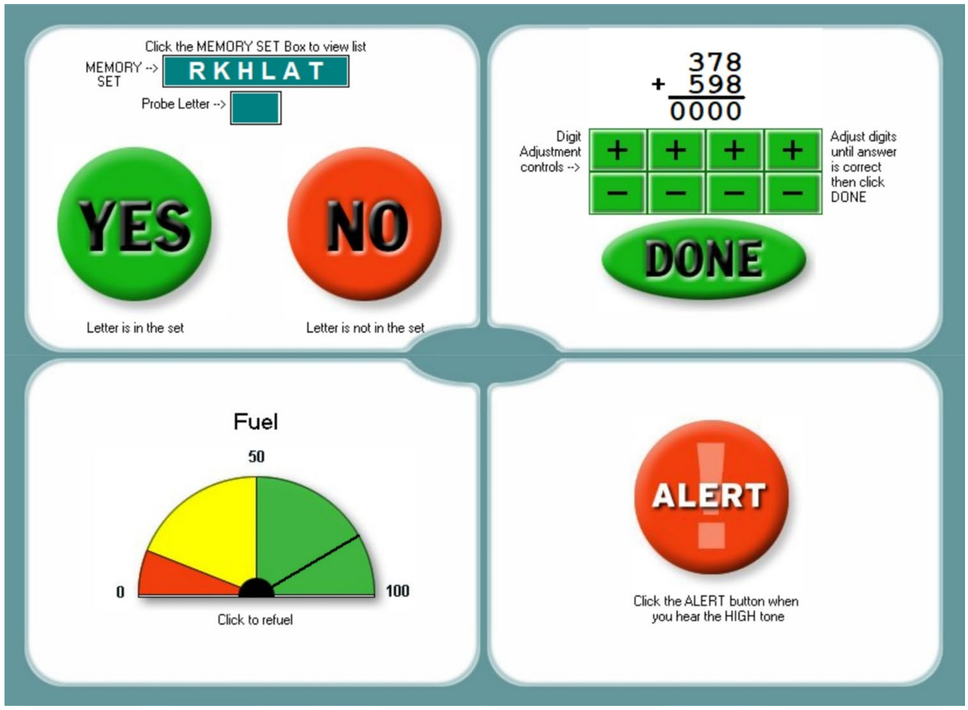

Fig. 2 SynWin. The four subtasks are memory search (top left); math (top right); visual monitoring (bottom left); and auditory monitoring (bottom right)

exclusive, in that they are collected on different sets of trials (error trials vs. correct trials). Empirically, both measures are higher on trials immediately following interruptions than on other trials, but we collapse across trial type because the same placekeeping mechanisms are hypothetically involved in both types of trials (Altmann & Trafton, 2015). The two measures also correlate strongly across the trial types, as we show in the Appendix.

In our sample, placekeeping error rates varied substantially across participants (0–81.4%, $M=10.6\%$, $SD=13.8\%$), as did response time (1.3–7.2 s, $M=3.5$ s, $SD=0.8$ s). The coefficient alpha, based on average performance on each block, was .88 for placekeeping errors and .83 for response time.

Multitasking

We used Elsmore's (1994) SynWin program to measure multitasking performance. Participants interleaved four subtasks, which are illustrated in Fig. 2. In the memory search subtask, a list of six letters was presented and then disappeared. A series of probe letters was then displayed. Participants determined whether each probe letter came from the original list or not; they were awarded 10 points for correct responses and penalized 10 points for incorrect responses or failures to respond. In the math subtask, participants added two three-digit numbers and clicked a "Done" button to register their answer. Participants were awarded 10 points for correct responses and penalized 10 points for incorrect responses. They had 20 s to answer each problem; every second after 20 s, 10 points were subtracted from their total score. In the visual monitoring subtask, participants clicked a "fuel gauge" before the indicator on the gauge reached zero; they earned more points for clicking the fuel gauge when the indicator was near zero, but lost 10 points for every second that the indicator was at zero. In the auditory monitoring subtask, participants responded to a high-pitched tone and ignored a low-pitched tone; they were awarded 10 points for correct responses and penalized 10 points for incorrect responses or failures to respond.

Participants completed one practice block and four test blocks of SynWin. Each block lasted 5 min. Scores from the four test blocks were averaged to create an overall multitasking score. The coefficient alpha for SynWin, based on average performance on each block, was .90.

Operation span

Participants solved math equations and remembered a letter that followed each equation (Unsworth, Heitz, Schrock, & Engle, 2005). After a series of trials, participants recalled the letters in the presented order, using the mouse to select each letter. There were 15 sets of equation-letter trials, 3 at each set size. Set size ranged from 3 to 7. The measure was the number of letters recalled in the correct order. The coefficient alpha for Operation Span, based on average performance at each set size, was .76.

Symmetry span

Participants made symmetry judgements about patterns and remembered the location of a square that appeared after each pattern (Unsworth et al., 2005). After a series of trials, participants recalled the location of the squares in the presented order, using the mouse to select each square location. There were 12 sets of pattern-square trials, 3 at each set size. Set size ranged from 2 to 5. The measure was the number of square locations recalled in the correct order. The coefficient alpha for symmetry span, based on average performance at each set size, was .66.

Raven's progressive matrices

Participants were presented with arrays of geometric patterns. Each array contained a missing item, and participants used the mouse to select a pattern that best completed the array. Participants were given 10 min to complete the 18 odd-numbered items from Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998). The measure was the number correct. The coefficient alpha, based on item-level performance, was .66.

Letter sets

Participants were presented with five sets of four letters arranged in a row, and used the mouse to select the set that did not follow the same pattern as the other four (Ekstrom, French, Harman, & Dermen, 1976). For example, for the sets NLIK, PLIK, QLIK, THIK, and VLIK, the correct response is THIK because the other sets all contain L. Participants were given 5 min to complete 20 items. The measure was the number correct. The coefficient alpha, based on item-level performance, was .70.

Number series

Participants were presented with a series of numbers, and used the mouse to select which of four alternatives logically completed the series. Participants were given 4.5 min to complete 15 items from the test of primary mental abilities (Thurstone, 1938). The measure was the number correct. The coefficient alpha, based on item-level performance, was .69.

Pattern comparison

Participants were presented with two patterns (i.e., symbols made from simple line drawings) on either side of a line and used the keyboard to indicate whether they were the same or different. Participants were given 30 s per set of 30 items; there were 2 sets of items (Salthouse & Babcock, 1991). The measure was the number correct minus two times the number incorrect. The coefficient alpha, based on average performance on each set, was .67.

Letter comparison

Participants were presented with sets of 3, 6, or 9 consonants on either side of a line and used the keyboard to indicate whether they were the same or different. Participants were given 30 s per set of 54 items; there were 2 sets of items (Salthouse & Babcock, 1991). The measure was the number correct minus the number incorrect. The coefficient alpha, based on average performance on each set, was .74.

Number comparison

Participants were presented with sets of 3, 6, or 9 numbers on either side of a line and used the keyboard to indicate whether they were the same or different. Participants were given 30 s per set of 54 items; there were 2 sets of items (Salthouse & Babcock, 1991). The measure was the number correct minus the number incorrect. The coefficient alpha, based on average performance on each set, was .81.

Vocabulary (synonyms and antonyms)

Participants were presented with a target word and four words that served as response options. For synonym items, participants used the mouse to click the response option most similar in meaning to the target word. For antonym items, participants used the mouse to click the response option most nearly the opposite in meaning to the target word. Participants were given 5 min for 10 synonym items and 5 min for 10 antonym items (Hambrick, Salthouse, & Meinz, 1999). The measure for each was the number correct. The coefficient alpha, based on item-level performance, was .37 for synonyms and .36 for antonyms.

Reading comprehension

Participants were presented with 10 items taken from an Armed Services Vocational Aptitude Battery practice test (Wiener & Steinberg, 1997). Each item consisted of a short paragraph and a related multiple-choice question, to which the participant responded using the keyboard. There was a time limit of 8 min, and the measure was the number correct. The coefficient alpha, based on item-level performance, was .40.

Results

Descriptive statistics for the cognitive ability measures are presented in Table 1 and correlations in Table 2. In general, correlations between placekeeping ability and

Table 1 Descriptive statistics for placekeeping ability and other cognitive ability measures

Measure	*N*	*M*	*SD*	Skewness	Kurtosis	α
SynWin multitasking	404	639.74	184.01	– 0.42	0.87	.90
UNRAVEL placekeeping ER (%)	404	10.58	13.80	2.91	9.23	.88
UNRAVEL response time (s)	404	3.50	0.81	1.03	1.85	.83
Operation span	399	38.49	18.18	– 0.16	– 0.63	.76
Symmetry span	401	18.01	9.07	0.25	– 0.60	.66
Raven's matrices	403	8.90	2.97	– 0.02	– 0.38	.66
Letter sets	403	10.18	2.88	0.08	– 0.44	.70
Number series	402	9.07	2.72	– 0.17	– 0.47	.69
Pattern comparison	403	37.14	8.76	– 0.42	0.45	.67
Letter comparison	401	20.28	4.09	0.02	– 0.52	.74
Number comparison	401	31.33	4.73	– 0.73	1.08	.81
Synonyms	401	3.20	1.78	0.38	– 0.31	.37
Antonyms	401	3.57	1.79	0.43	0.03	.36
Reading comprehension	402	8.58	1.34	– 0.97	0.46	.40

Note. ER, error rate

Table 2 Correlations for multitasking, placekeeping ability, and cognitive ability measures

	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1. SynWin multitasking	–													
2. UNRAVEL placekeeping ER	**– .25**	–												
3. UNRAVEL response time	**– .29**	**.33**	–											
4. Operation span	**.38**	**– .25**	**– .30**	–										
5. Symmetry span	**.38**	**– .15**	**– .23**	**.42**	–									
6. Raven's matrices	**.41**	**– .24**	– .10	**.12**	**.19**	–								
7. Letter sets	**.42**	**– .19**	**– .29**	**.30**	**.21**	**.23**	–							
8. Number series	**.51**	**– .22**	**– .23**	**.20**	**.20**	**.38**	**.32**	–						
9. Pattern comparison	**.35**	– .08	**– .14**	**.10**	**.24**	**.28**	**.18**	**.22**	–					
10. Letter comparison	**.25**	**– .12**	**– .26**	**.16**	**.14**	.06	**.27**	**.18**	**.25**	–				
11. Number comparison	**.36**	– .08	**– .28**	.08	**.13**	.08	**.25**	**.23**	**.28**	**.57**	–			
12. Synonyms	**.14**	**– .17**	– .08	**.10**	.07	**.25**	.05	**.22**	**.12**	.08	**.11**	–		
13. Antonyms	**.19**	**– .14**	– .08	**.12**	.02	**.23**	**.11**	**.18**	.06	**.10**	.09	**.36**	–	
14. Reading comprehension	**.24**	**– .13**	– .08	.03	.09	**.27**	.07	**.27**	**.14**	**.14**	.07	**.27**	**.28**	–

Note. Listwise $N = 391$. ER, error rate. Bold, $p < .05$

the other measures were negative and significant, indicating better placekeeping performance (fewer placekeeping errors, faster response times) for higher ability participants. The average correlation between the two placekeeping measures (placekeeping errors and response time) and multitasking performance was $r = .27$, in line with correlations of $r = .20$–$.40$ between cognitive abilities and multitasking performance in previous work (see, e.g., König, Buhner, & Murling, 2005). Placekeeping correlated less strongly with multitasking than did other measures (e.g., measures of fluid intelligence; $.41 \leq r \leq .51$), but this comparison leaves open the central question of whether variance is shared across predictor variables, which we address in the following sections.

Predictive validity

We first performed regression analyses to estimate the proportion of variance in multitasking performance accounted for by placekeeping ability. The results are presented in Table 3. In the first analysis, we aggregated the placekeeping measures across the full UNRAVEL session. In four subsequent analyses, we separated the placekeeping measures by test block to assess their stability.

In the full-session analysis, both placekeeping measures accounted for a significant amount of variance in multitasking performance (ps < .001), and the full model accounted for 11% of the variance in multitasking performance, $F(2, 401) = 25.53$, $p < .001$. Thus, higher levels of placekeeping

Table 3 Regression analyses testing predictive validity of placekeeping ability for multitasking performance

Measure	B	SE_B	β	t	F	p	R^2
Full session							
Placekeeping error rate	− 222.06	66.56	− .167	− 3.34		< .001	
Response time	− 54.66	11.31	− .241	− 4.83		< .001	
Full model					25.53	< .001	.113
Block 1							
Placekeeping error rate	− 228.37	49.39	− .223	− 4.62		< .001	
Response time	− 19.13	6.83	− .135	− 2.80		.005	
Full model					15.19	< .001	.070
Block 2							
Placekeeping error rate	− 221.16	64.78	− .167	− 3.41		< .001	
Response time	− 41.57	8.91	− .229	− 4.67		< .001	
Full model					21.86	< .001	.099
Block 3							
Placekeeping error rate	− 164.64	64.27	− .126	− 2.56		.011	
Response time	− 45.91	10.36	− .218	− 4.43		< .001	
Full model					16.42	< .001	.076
Block 4							
Placekeeping error rate	− 223.90	53.46	− .197	− 4.19		< .001	
Response time	− 64.47	11.48	− .264	− 5.62		< .001	
Full model					26.19	< .001	.116

Note. Full model $df_1 = 2$ and $df_2 = 401$ except for Block 2. For Block 2, full model $df_1 = 2$ and $df_2 = 399$; missing cases arose when participants had no correct trials in a block on which to measure response time

ability were associated with better multitasking. The block-wise analyses showed the same pattern. Both placekeeping measures accounted for a significant proportion of variance in multitasking in each block ($ps \leq .011$), indicating that the predictive validity of placekeeping for multitasking was stable across samples of performance and robust to placekeeping practice.

Incremental validity

With a relationship established between placekeeping and multitasking, we conducted a series of hierarchical regression analyses to test whether placekeeping ability contributed to multitasking performance above and beyond other cognitive ability measures. Each analysis involved two steps. In Step 1, we added either a cognitive ability measure or a composite variable formed from a set of cognitive ability measures. In Step 2, we added placekeeping error rate and response time. Composite variables were formed by averaging z scores (i.e., standardized scores) for the constituent measures. There were four composite variables, each representing a broad cognitive ability: working memory capacity (measures: symmetry span, operation span), fluid intelligence (measures: Raven's matrices, letter sets, number series), perceptual speed (measures: pattern comparison, letter comparison, number comparison), and crystallized intelligence (measures: synonyms, antonyms, reading comprehension).

As a final analysis, we entered psychometric g in Step 1, and entered placekeeping ability in Step 2. Psychometric g was estimated by saving scores on the first unrotated component from a principal components analysis of the 11 cognitive ability measures (Jensen, 2002). A set of 10 or more cognitive ability measures, if they measure diverse abilities, generally produces stable estimates of g (Reeve & Blacksmith, 2009).

The results are presented in Table 4. The cognitive ability measures and composite variables each accounted for a significant amount of variance in multitasking performance, in amounts similar to those in other studies of the relationship between cognitive ability and multitasking performance (e.g., Hambrick et al., 2011). Addressing our research question, placekeeping contributed significantly to the prediction of multitasking performance above and beyond each of the other individual cognitive ability measures and each of the composite variables. Placekeeping accounted for 3.9% of the variance in multitasking performance above and beyond working memory capacity, 2.0% of the variance above and beyond fluid intelligence, 5.5% of the variance above and beyond perceptual speed, and 8.5% of the variance above and beyond crystallized intelligence. Placekeeping did not contribute significantly to the prediction of multitasking performance above and beyond psychometric g, a point we return to in the Discussion.

Table 4 Hierarchical regression analyses testing incremental validity of placekeeping ability for multitasking performance

Measure or composite	Step 1: measure or composite entered R^2	df_2	p	Step 2 (incremental validity): placekeeping measures entered ΔR^2	df_2	p	Total R^2
Operation span	.140	397	< .001	.049	395	< .001	.189
Symmetry span	.143	399	< .001	.064	397	< .001	.207
WMC composite	.195	401	< .001	.039	399	< .001	.235
Raven's matrices	.164	401	< .001	.072	399	< .001	.237
Letter sets	.164	401	< .001	.052	399	< .001	.216
Number series	.264	400	< .001	.043	398	< .001	.306
Gf composite	.363	401	< .001	.020	399	.002	.383
Pattern comparison	.124	401	< .001	.082	399	< .001	.206
Letter comparison	.056	399	< .001	.085	397	< .001	.141
Number comparison	.128	399	< .001	.070	397	< .001	.198
PS composite	.168	401	< .001	.055	399	< .001	.223
Synonyms	.025	399	.001	.099	397	< .001	.124
Antonyms	.033	399	< .001	.098	397	< .001	.131
Reading comprehension	.061	400	< .001	.094	398	< .001	.154
Gc composite	.070	401	< .001	.085	399	< .001	.156
Psychometric *g*	.437	389	< .001	.004	387	.287	.441

Note. WMC, working memory capacity; Gf, fluid intelligence; PS, perceptual speed; Gc, crystallized intelligence. For all models, Step 1 $df_1 = 1$ and Step 2 $df_1 = 2$

Discussion

Multitasking often involves performance of procedures interleaved with each other or with other tasks, in task environments where failing to remember where one left off on a given subtask or what one was planning to do next could result in mistakes or lost time. Therefore, we asked whether measures of placekeeping ability could help explain variance in multitasking performance. We also asked whether placekeeping could explain variance in multitasking beyond each of a diverse and widely studied set of cognitive ability measures, to develop evidence on whether placekeeping is a distinct cognitive ability.

We found that placekeeping ability accounted for a significant amount of the variance in multitasking performance ($R^2 = 11\%$), and that the relationship was stable across blocks of placekeeping performance (Table 3), indicating that the predictive validity is robust to effects of placekeeping practice. We also found that placekeeping ability accounted for a significant amount of the variance in multitasking performance above and beyond each of the other cognitive ability measures we tested (Table 4). The diversity of the measures reduces the likelihood that placekeeping is subsumed in an ability that we did not test—although other abilities remain to be tested, a point we return to below. Thus, one interpretation of our results is that placekeeping is a distinct cognitive ability.

Two of the abilities we tested were of particular interest because there are reasons to think the variance they explained in our outcome measure would overlap completely with that explained by placekeeping. One was working memory capacity, which overlaps with placekeeping in terms of the need to remember recent events. However, placekeeping also hypothetically depends on long-term memory to store knowledge of the steps and their correct sequence (Altmann & Trafton, 2015). In context of multitasking, these long-term representations might include procedures for individual subtasks and a strategy for interleaving them. Consistent with this analysis, we found that placekeeping accounted for variance in our criterion task that working memory capacity did not. This result converges with a previous finding—that placekeeping had incremental validity for fluid intelligence relative to working memory capacity (Burgoyne et al., 2019)—to suggest that placekeeping and working memory capacity are distinct.

The other ability of particular interest was fluid intelligence. Of the broad cognitive abilities, this is the one most likely to explain all the variance explained by any other ability one might choose to measure (see, e.g., Carpenter, Just, & Shell, 1990). Nonetheless, placekeeping captured a small but significant amount of variance in multitasking that fluid intelligence did not. The amount (2.0%) was smaller than for any of the other comparison abilities (3.9–8.5%), suggesting that placekeeping is in a sense most similar to fluid intelligence. Converging evidence for this similarity is that placekeeping has predictive validity for fluid intelligence (Hambrick & Altmann, 2015) and incremental validity for fluid intelligence relative to working memory capacity (Burgoyne et al., 2019). One interpretation of this similarity is that placekeeping can be conceptualized as a component of fluid intelligence that is not fully represented by existing indicators such as those we used in this study. Fluid intelligence is the ability to solve novel problems, and task analysis suggests that problem solving depends heavily on placekeeping to support efficient problem-space search (Hambrick et al., 2020). Arguably, then, placekeeping is less a distinct cognitive ability than a core component of fluid intelligence, and our results can be taken to mean that a placekeeping measure might be a useful addition to existing indicators of fluid intelligence.

Placekeeping did not show incremental validity for multitasking above and beyond psychometric *g* (Table 4). Psychometric *g* typically accounts for as much criterion variance as any specific aptitude does (Schmidt, Ones, & Hunter, 1992), so this finding places placekeeping at the level of a specific aptitude rather than a broad ability. This level seems consistent with our view that placekeeping represents a specific form of cognitive control that plays a supporting role in overall task performance (Altmann et al., 2014), even if it is general in the sense that it plays a role in many tasks.

Finally, we assessed the potential of the UNRAVEL task to predict multitasking performance when measures beyond placekeeping were included as predictors. UNRAVEL is a complex task, and as such may predict criterion variance based on the involvement of a variety of cognitive processes. For applications such as personnel selection, where testing time is limited or costly, it makes sense to make the most of each instrument. Accordingly, we performed an analysis aimed at assessing the maximum predictive validity of UNRAVEL for multitasking. The details are summarized here and reported in full in the Appendix. Four measures were predictive. Two of these were placekeeping measures, and the other two were the duration of the introductory phase at the start of an UNRAVEL session, which measures time taken to acquire the task, and the duration of interruptions, which measures typing speed and accuracy. These four measures explained 1.6% of variance in multitasking above and beyond psychometric *g* ($p = .022$ based on the full session of UNRAVEL performance and $\Delta R^2 = 1.5\%$, $p = .033$ based on just the first test block). This incremental validity with respect to *g* suggests that UNRAVEL could add predictive value to specific aptitude tests, which, as we noted earlier, are unlikely to explain more criterion variance than *g*.

In terms of predictive validity, our all-measures model explained 21% of the variance in multitasking (Table 7), which is nearly double that explained by the placekeeping-only model (11%; Table 3) but still substantially less than the 36% explained by fluid intelligence (Table 4). That said, fluid intelligence is often tested using tasks that may have a liability, which is that they are puzzle tasks (like Raven's Matrices) that can cease to be novel problems after they are solved once. Thus, learning by the performer during the first administration may undermine the construct validity of future administrations. Although we know of no direct evidence on this issue, there is evidence that repeated testing on Raven's Matrices leads to changes in test-taking strategies, with the changes accounting for substantial variance in score gains across administrations (Hayes, Petrov, & Sederberg, 2015). In other words, the abilities measured by Raven's Matrices and other puzzle tests may change with repeated testing as learners acquire skills specific to the task.

Relative to this potential concern, we found that the incremental validity of UNRAVEL for multitasking was stable across blocks (Table 3). Similarly, Burgoyne et al. (2019) found that different blocks loaded equally highly on a latent placekeeping factor. We also reanalyzed the data from Hambrick and Altmann (2015), and found that the relationship between placekeeping error rate and fluid intelligence remained unchanged across administrations of UNRAVEL on separate days (*r*s = − .41 and − .40, test of difference $z = 0.15$, $p = .879$). Thus, the UNRAVEL task appears to be "reusable" as opposed to "one-shot," and when the testing history of job candidates is unknown, a collection of such reusable tasks, each with incremental validity relative to the others, may be preferable to a battery that focuses on solving problems that can be learned.

In terms of future work, we suggest two directions. One is to test the incremental validity of placekeeping relative to additional broad cognitive abilities, to refine our understanding of whether placekeeping is a distinct construct. The four comparison abilities we tested here are widely studied and diverse, but others could be considered. Of particular interest is broad retrieval ability, which addresses memory storage, consolidation, and fluent retrieval (McGrew, 2009). In conjunction with working memory capacity, broad retrieval ability could subsume the memory operations required for placekeeping (Altmann & Trafton, 2015) and thus the construct as a whole. A second direction for future research would be to develop and validate additional tests of placekeeping ability and multitasking performance so that analyses can be shifted from the level of observed variables to the level of latent variables, which are free of measurement error and closer to the theoretical constructs of interest.

Funding This study was funded by Grants from the Office of Naval Research (N00014-13-1-0247, N00014-16-1-2457, and N00014-16-1-2841).

Compliance with ethical standards

Conflict of interest All authors declare that they have no conflict of interest.

Ethical approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. This article does not contain any studies with animals performed by any of the authors.

Informed consent Informed consent was obtained from all individual participants included in the study.

Open practice statement The data for this study will be made available upon request.

Appendix

Here we explore the predictive and incremental validity of the UNRAVEL task as a whole, looking beyond placekeeping measures. Like many complex tasks, UNRAVEL affords multiple measures of behavior that collectively reflect a range of different cognitive processes. From a practical perspective, when testing time is costly or limited it makes sense to wring the most possible from each instrument in a battery, and we wanted to assess what predictive value the UNRAVEL task would have for multitasking if we used every predictive piece of it.

We tested seven measures selected to span all aspects of task performance during an UNRAVEL session. The measures are exhaustive but also exclusive, in that they are recorded from non-overlapping sets of events (e.g., from distinct sets of trials). Descriptive statistics and reliabilities are presented in Table 5. Six of the seven measures had acceptable reliability (coefficient alpha > .70). The seventh (introduction duration, described below) did not have repeated measures from which to compute reliability. Correlations among the measures and between the measures and multitasking are presented in Table 6. Six of seven measures correlated significantly with multitasking in the expected (negative) direction, and the seventh trended in that direction.

We tested three new measures: the choice-rule error rate, introduction duration, and interruption duration. The *choice-rule error rate* is the proportion of trials on which the participant selects the correct step but chooses the wrong response according to the rule for that step (see Fig. 1). Choice-rule errors and placekeeping errors thus occur on different trials, and neither is included in response time, which is recorded on correct trials only. Together, choice-rule errors, placekeeping errors, and response times measure behavior on all trials during test blocks. *Introduction duration* is the time spent on the introductory phase that occurs before the test blocks in an experimental session. During this phase, participants learn the task and receive practice trials and interruptions. Introduction duration incorporates effects of errors, because the computer requires correct responses during this phase, so an error costs time as the participant must respond again. Finally, *interruption duration* is the time spent typing the "codes" presented during interruptions in test blocks. This measure similarly incorporates effects of errors, because an incorrectly typed code costs time as the participant must try again.

The four remaining measures were related to placekeeping. Here we divided the placekeeping error rate and response time each into two variants according to a distinction between *post-interruption trials*, which immediately follow an interruption, and *non-interruption trials*, which immediately follow another trial. Post-interruption trials generate more placekeeping errors and longer response times than non-interruption trials, reflecting the effect of interruptions on memory for the most recently performed step (Altmann et al., 2017). In the body of the paper, we combined the two trial types because the underlying placekeeping operations are hypothetically the same, differing only in the age of memory for recently performed trials (Altmann & Trafton, 2015). Here we separate them to explore the effect of their empirical differences on predictive validity. We do not separate choice-rule errors by trial

Table 5 Descriptive statistics for expanded set of UNRAVEL measures

Measure	*M*	*SD*	Skewness	Kurtosis	α
Choice-rule error rate (%)	2.06	2.80	4.83	39.70	.85
Introduction duration (s)	517.44	157.91	1.76	4.91	n/a
Interruption duration (s)	21.64	6.52	0.77	1.27	.96
PI placekeeping error rate (%)	23.05	19.94	1.37	1.67	.85
NI placekeeping error rate (%)	8.07	13.47	3.24	11.27	.87
PI response time (s)	4.92	1.61	2.49	17.68	.71
NI response time (s)	3.28	0.79	1.11	2.30	.84

Note. PI, post-interruption; NI, non-interruption. N = 404. α is computed from average performance on each test block; no repeated measures were available to compute α for introduction duration

Table 6 Correlations for multitasking and expanded set of UNRAVEL measures

	1	2	3	4	5	6	7	8
1. SynWin multitasking	–							
2. Choice-rule error rate	− .09	–						
3. Introduction duration	**− .29**	.02	–					
4. Interruption duration	**− .30**	.04	**.32**	–				
5. PI placekeeping error rate	**− .30**	**.32**	**.21**	**.10**	–			
6. NI placekeeping error rate	**− .22**	**.21**	**.15**	**.12**	**.71**	–		
7. PI response time	**− .24**	**.10**	**.13**	**.22**	**.34**	**.25**	–	
8. NI response time	**− .29**	.00	**.27**	**.31**	**.27**	**.33**	**.56**	–

Note. PI, post-interruption; NI, non-interruption. Listwise $N=404$. Bold, $p < .05$

Table 7 Regression analysis testing predictive validity of UNRAVEL measures for multitasking performance

Measure	B	SE_B	β	t	F	p	R^2
Choice-rule error rate	− 45.90	312.26	− .007	− 0.15		.883	
Introduction duration	− 0.18	0.06	− .152	− 3.11		**.002**	
Interruption duration	− 5.18	1.38	− .183	− 3.74		**< .001**	
PI placekeeping error rate	− 218.58	63.33	− .237	− 3.45		**< .001**	
NI placekeeping error rate	69.10	89.66	.051	0.77		.441	
PI response time	− 4.14	6.42	− .036	− 0.64		.520	
NI response time	− 29.97	13.67	− .128	− 2.19		**.029**	
Full model					14.64	**< .001**	.206

Note. PI, post-interruption; NI, non-interruption. Bold, $p < .05$. Full model $df_1=7$ and $df_2=396$

type, because empirically they do not differ by trial type (Altmann et al., 2014).

Predictive validity

We performed a regression analysis to test the predictive validity of these seven measures for multitasking performance. The results are presented in Table 7.

There are two main findings. First, the full model explained 21% of the variance in multitasking performance. This is nearly double the 11% of variance explained when we restricted our analysis to placekeeping measures (Table 3), indicating that there is more to UNRAVEL performance than placekeeping, and that some of the additional processes also play a role in multitasking. Twenty-one percent trends larger than the variance explained by working memory capacity and perceptual speed (20% and 17%, respectively; Table 4), and is substantially more than the variance explained by crystallized intelligence (7%), but is substantially less than the variance explained by fluid intelligence (36%).

Second, four measures—introduction duration, interruption duration, the post-interruption placekeeping error rate, and non-interruption response time—were significant predictors, suggesting that they captured systematic variance that is worth trying to interpret in theoretical terms. Introduction duration may measure a person's ability to acquire a procedure quickly, and for multitasking may predict the ability to develop procedures or strategies for interleaving subtasks. Interruption duration measures speed and accuracy of typing and presumably of the perceptual processes needed to correctly encode a string of randomly-ordered letters, processes that seem basic to many tasks. Interruption duration is also a predictor of the placekeeping error rate in our cognitive model (Altmann, et al., 2017) because it affects memory for the step performed before the interruption, but any placekeeping-related variance it explains in multitasking may be mediated by the placekeeping error rate.

Finally, the post-interruption placekeeping error rate and non-interruption response time are two of the four variants of our placekeeping measures. The other two variants—the non-interruption placekeeping error rate and post-interruption response time—were not significant predictors. In statistical terms, the reason is probably that, with reference to Table 6, each non-significant measure correlated strongly with its significant counterpart (e.g., $r=.71$ for non-interruption and post-interruption placekeeping errors), but correlated somewhat less strongly with multitasking than did its counterpart (e.g., $r=-.22$ for non-interruption placekeeping errors and multitasking vs. $r=-.30$ for post-interruption placekeeping errors and multitasking) and thus was dominated in the model by its significant counterpart. The strong correlations between the two placekeeping error variants and between the two response time variants support our assumption earlier that the same underlying mechanisms account for performance of the two trial types. Whether the somewhat different correlations with multitasking have a theoretical basis or merely reflect sampling variability is an interesting question for future work.

Incremental validity

We performed hierarchical regression analyses to test the incremental validity of the four significant UNRAVEL predictors relative to the measure of psychometric *g* we reported in the body of the paper. In Step 1, we entered psychometric *g*, and in Step 2, we entered the four significant UNRAVEL predictors. We performed this analysis with the UNRAVEL measures aggregated over the full session and also for Block 1 alone to assess the possibility of using a reduced version of the task to save testing time.

The UNRAVEL measures entered in Step 2 explained an additional 1.6% of variance in multitasking performance, $F(4, 385) = 2.90$, $p = .022$ for the full-session analysis and an additional 1.5% of variance, $F(4, 385) = 2.66$, $p = .033$ for the Block 1 analysis. Thus, an expanded set of UNRAVEL measures showed incremental validity for multitasking relative to psychometric *g*, whether they were collected from the whole session or only from Block 1. In terms of testing time, there would be considerable savings from using just Block 1. The duration of the introductory phase plus Block 1 ($M = 18$ m, $SD = 3.8$ m) was substantially shorter than the duration of the full session ($M = 43$ m, $SD = 8.5$ m), and is similar to the total duration of the three fluid intelligence indicators (about 20 m).

An additional 1.6% of criterion variance explained may seem a trivial amount, but increments in R^2 underestimate the practical significance of incremental validity (Taylor & Russell, 1939; see also Hambrick, Burgoyne, & Oswald, 2019). For example, per Taylor and Russell's tables, an additional 1.0% of criterion variance explained (i.e., $r = .10$) increases the proportion of new employees considered satisfactory by 5%, for a job in which 50% of existing employees are considered satisfactory and the selection ratio for new employees is 30%. For jobs that require extensive training, or in which procedural or placekeeping errors are especially costly, the benefits of the extra test could outweigh the costs. For purposes of this analysis, we suppose that if a given instrument—UNRAVEL, here, with all significant predictors used—explains criterion variance above and beyond psychometric *g*, it may also explain criterion variance above and beyond any individual aptitude test, given that specific aptitudes rarely explain more variance than *g* (Schmidt et al., 1992). In general, other complex tasks may add predictive value also if task performance is measured in a similarly comprehensive manner.

References

Altmann, E. M., & Hambrick, D. Z. (2017). Practice increases procedural errors after task interruption. *Journal of Experimental Psychology: General, 146*, 615–620.

Altmann, E. M., & Trafton, J. G. (2015). Brief lags in interrupted sequential performance: Evaluating a model and model evaluation method. *International Journal of Human-Computer Studies, 79*, 51–65.

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2014). Momentary interruptions can derail the train of thought. *Journal of Experimental Psychology: General, 143*, 215–226.

Altmann, E. M., Trafton, J. G., & Hambrick, D. Z. (2017). Effects of interruption length on procedural errors. *Journal of Experimental Psychology: Applied, 23*, 216–229.

Burgoyne, A. P., Hambrick, D. Z., & Altmann, E. M. (2019). Placekeeping ability as a component of fluid intelligence: Not just working memory capacity. *American Journal of Psychology, 132*, 439–449.

Carlson, R. A., & Cassenti, D. N. (2004). Intentional control of event counting. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 30*, 1235–1251.

Carpenter, P. A., Just, M. A., & Shell, P. (1990). What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test. *Psychological Review, 97*, 404–431.

Cattell, R. B. (1943). The measurement of adult intelligence. *Psychological Bulletin, 40*, 153–193.

Ekstrom, R., French, J., Harman, H., & Dermen, D. (1976). *Kit of factor-referenced cognitive tests* (Rev ed.). Princeton, NJ: Educational Testing Service.

Elsmore, T. F. (1994). SYNWORK1: A PC-based tool for assessment of performance in a simulated work environment. *Behavior Research Methods, Instruments, and Computers, 26*, 421–426.

Hambrick, D. Z., & Altmann, E. M. (2015). The role of placekeeping ability in fluid intelligence. *Psychonomic Bulletin and Review, 22*, 1104–1110.

Hambrick, D. Z., Altmann, E. M., & Burgoyne, A. P. (2018). A knowledge-activation approach to testing the Circumvention-of-Limits hypothesis. *American Journal of Psychology, 131*, 307–321.

Hambrick, D. Z., Burgoyne, A. P., & Altmann, E. M. (2020). Intelligence and problem solving. In: R. J. Sternberg (Ed.), *The Cambridge Handbook of Intelligence*, 2nd edition. Cambridge: Cambridge University Press.

Hambrick, D. Z., Burgoyne, A. P., & Oswald, F. L. (2019). Domain-general models of expertise: The role of cognitive ability. In P. Ward, J. M. Schraagen, J. Gore, & E. Roth (Eds.), *The Oxford Handbook of Expertise*. New York: Oxford University Press.

Hambrick, D. Z., Oswald, F. L., Darowski, E. S., Rench, T. A., & Brou, R. (2010). Predictors of multitasking performance in a synthetic work paradigm. *Applied Cognitive Psychology, 24*, 1149–1167.

Hambrick, D. Z., Salthouse, T. A., & Meinz, E. J. (1999). Predictors of crossword puzzle proficiency and moderators of age–cognition relations. *Journal of Experimental Psychology: General, 128*, 131–164.

Hambrick, D. Z., Rench, T. A., Poposki, E. M., Darowski, E. S., Roland, D., Bearden, R. M., et al. (2011). The relationship between the ASVAB and multitasking in Navy Sailors: A process-specific approach. *Military Psychology, 23*, 365–380.

Hayes, T. R., Petrov, A. A., & Sederberg, P. B. (2015). Do we really become smarter when our fluid-intelligence test scores improve? *Intelligence, 48*, 1–14.

Jensen, A. R. (2002). Psychometric *g*: Definition and substantiation. In *The General Factor of Intelligence* (pp. 51–66). Psychology Press.

König, C. J., Buhner, M., & Murling, G. (2005). Working memory, fluid intelligence, and attention are predictors of multitasking performance, but polychronicity and extraversion are not. *Human Performance, 18*, 243–266.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence, 37*, 1–10.

National Center for O*NET Development (2019). Browse by O*NET Data. O*NET OnLine. Retrieved December 22, 2019, from https://www.onetonline.org/find/descriptor/result/1.A.1.g.2.

Raven, J., & Raven, J. C., & Court, J. H. (1998). *Manual for Raven's Progressive Matrices and Vocabulary Scales*.

Reeve, C. L., & Blacksmith, N. (2009). Identifying g: A review of current factor analytic practices in the science of mental abilities. *Intelligence, 37*, 487–494.

Salthouse, T. A., & Babcock, R. L. (1991). Decomposing adult age differences in working memory. *Developmental Psychology, 27*, 763–776.

Schmidt, F. L., Ones, D. S., & Hunter, J. E. (1992). Personnel selection. *Annual Review of Psychology, 43*, 627–670.

Taylor, H. C., & Russell, J. T. (1939). The relationship of validity coefficients to the practical effectiveness of tests in selection: Discussion and tables. *Journal of Applied Psychology, 23*, 565–578.

Thurstone, L. L. (1938). Primary mental abilities. *Psychometric Monographs, 1*, 270–275.

Unsworth, N., Heitz, R. P., Schrock, J. C., & Engle, R. W. (2005). An automated version of the operation span task. *Behavior Research Methods, 37*, 498–505.

Wang, D. Y. D., Proctor, R. W., & Pick, D. F. (2007). Acquisition and transfer of attention allocation strategies in a multiple-task work environment. *Human Factors, 49*, 995–1004.

Wiener, S., & Steinberg, E. P. (1997). *Everything you need to score high on the ASVAB* (15th ed.). New York: Macmillan.