A Knowledge Activation Approach to Testing the Circumvention-of-Limits Hypothesis

Author(s): David Z. Hambrick, Erik M. Altmann and Alexander P. Burgoyne

Source: *The American Journal of Psychology*, Vol. 131, No. 3 (Fall 2018), pp. 307-321

Published by: University of Illinois Press

Stable URL: https://www.jstor.org/stable/10.5406/amerjpsyc.131.3.0307

A Knowledge Activation Approach to Testing the Circumvention-of-Limits Hypothesis

DAVID Z. HAMBRICK, ERIK M. ALTMANN, and ALEXANDER P. BURGOYNE
Michigan State University

The circumvention-of-limits hypothesis holds that the more skilled and knowledgeable the task performer, the less it matters for task performance whether that person has limited general cognitive ability. We tested this hypothesis using a knowledge activation approach to manipulate knowledge experimentally. The criterion task, which we designed to capture a fundamental requirement of a broad class of real-world tasks, was a placekeeping task in which participants had to perform a sequence of operations in a specified order, applying 7 different 2-alternative decision rules to a series of randomly generated stimuli. The measures of interest in this task were response time and accuracy, as well as frequency of use of a help function to recall the correct sequence of steps. In the knowledge-activated condition, we gave participants a mnemonic in which the first letters of the decision rules spelled the English word *unravel*. In the knowledge-not-activated condition, no mnemonic was given, and the use and discovery of mnemonics was frustrated by reversal of the terms of some decision rules so that their first letters spelled the difficult to pronounce nonword *unrbcel*. The predictor tasks, which we used to predict performance in the placekeeping task, were standard tests of cognitive ability, which have been shown to predict performance differences in a wide range of complex tasks. Inconsistent with the circumvention-of-limits hypothesis, the positive effect of cognitive ability on placekeeping performance did not differ across conditions. This finding adds to previous evidence that it may not always be possible to overcome limitations on cognitive ability through extensive training.

KEYWORDS: expertise, knowledge, cognitive ability, intelligence, placekeeping

The question of what role, if any, general cognitive ability plays in skilled performance is the subject of vigorous debate in the literature on expertise (see Hambrick et al., 2014). The basic idea of the circumvention-of-limits hypothesis (Hambrick & Meinz, 2011; Salthouse, 1991) is that the effect of domain-general abilities and capacities on performance diminishes as skill in a task is acquired through training. In other words, higher cognitive ability translates into superior performance initially but then is replaced by domain-specific knowledge as the most important predictor of performance. This idea, which follows from classic theories of skill acquisition (Fitts & Posner, 1967), was expressed by Ericsson and Charness (1994) as follows: "Performers can acquire skills that circumvent basic limits on working memory capacity and sequential processing" (p. 725). More recently, Ericsson (2014) stated that "acquired mechanisms gradually circumvent the role of any basic general cognitive capacities and thus reduce and even elimi-

American Journal of Psychology
Fall 2018, Vol. 131, No. 3 pp. 307–321 •

nate significant relations between general cognitive ability and domain-specific performance at the expert level of performance" (p. 83).

One of the challenges of testing the circumvention-of-limits hypothesis is that it is difficult to experimentally manipulate expertise in the laboratory. For example, it would take years to train participants who have never played chess to become chess masters, with no guarantee that all participants would reach that level of skill (Gobet & Campitelli, 2007). Consequently, the typical approach in research on expertise is to compare intact groups of novices and experts (e.g., Chase & Simon, 1973). However, a problem with using intact groups is that this approach is quasiexperimental and thus vulnerable to confounds. For example, age and expertise in some domains are highly confounded, and finding older novices or young experts in sufficient numbers is difficult. This problem is illustrated by a recent meta-analysis by Burgoyne et al. (2016) examining the relationship between cognitive ability and chess skill. Consistent with the circumvention-of-limits hypothesis, there was a stronger positive correlation between fluid intelligence and chess skill for samples that included unranked chess players than for samples that included ranked (and more skilled) chess players. However, the index of skill (unranked vs. ranked) was highly confounded with age, because almost all of the unranked samples were youths and almost all of the ranked samples were adults.

An alternative approach to testing the circumvention-of-limits hypothesis is to manipulate the activation of task-relevant knowledge experimentally in a laboratory task. The goal with this knowledge activation approach is to "add knowledge" to the mind of a participant in a tractably short amount of time before his or her performance is assessed. This addition is achieved through the use of instructions, administered before performance, that activate task-relevant knowledge that all participants can generally be expected to have and that aids task performance in some way. The advantage of this approach is that random assignment of participants to knowledge groups overcomes the internal validity problems that come with quasiexperimental studies of expert–novice differences. The challenge with this approach is to design a task that can be influenced by knowledge representations that are commonly found in the population and then find a way to activate those representations in one group of participants but not the other. A limitation of this approach is that the gap in knowledge (and performance differences) between the two groups will almost certainly be smaller than that created by extensive training, a point we return to in the *Discussion*.

In the study that introduced the knowledge activation approach (Hambrick & Oswald, 2005), participants performed a memory task in which they attempted to remember the movements of spaceships that "flew" from planet to planet in a solar system. Unbeknownst to participants, the spaceships flew in the same manner that baseball players run around a baseball diamond. Participants then performed an isomorphic task in which a baseball diamond replaced the solar system and baseball players replaced the spaceships. Finally, participants completed tests of working memory capacity and knowledge of baseball. The relationship between baseball knowledge and performance was greater in the baseball condition than in the spaceship condition, indicating activation of domain knowledge in the baseball condition. However, the relationship between working memory capacity and performance did not differ across conditions. That is, contrary to the circumvention-of-limits hypothesis, the relationship between working memory capacity and performance was as large in the baseball condition, where task-relevant knowledge was activated, as in the spaceship condition, where it was not.

Here, we extend the knowledge activation approach to placekeeping, a particular kind of "sequential processing" (Ericsson & Charness, 1994) necessary for performance in many task domains. Placekeeping involves carrying out a sequence of steps in a prescribed order, without skipping or repeating steps. Many everyday tasks include a placekeeping component, and placekeeping errors at many levels of seriousness are possible. In some cases, errors are merely irritating, as when one skips or repeats a step while following a recipe, spoiling the dish. In other cases, errors can be catastrophic, as when a pilot skips a critical step during a takeoff or landing procedure.

To study placekeeping in the laboratory, we developed a task in which there are seven decisions, or steps, to be performed in a fixed sequence (Altmann, Trafton, & Hambrick, 2014). Each step is performed on a randomly generated, multidimensional stimulus that affords performance of any of the seven steps. At the end of the sequence, the participant simply returns to the beginning of the sequence, generating a long continuous sequence of performance. Placekeeping is a challenge in this task because the stimulus affords no information about which step is currently correct, meaning that the participant has to remember which step is currently correct. Also, the task environment interrupts the participant at random points between steps with a simple transcription typing task, after which the participant must resume the sequence where he or she left off. Interruptions are common in many of the same real-world task environments in which placekeeping is required and have the effect of increasing the error rate at the point of task resumption. Interruptions thus increase the external validity of our task to some extent and also mitigate any problems with range restriction in the error rate. The task is designed to study memory processes, even though placekeeping is often supported by perceptual cues or "affordances"; for example, in the task of checking your mail, an open mailbox door cues the next step of reaching into the mailbox. However, even when perceptual placekeeping is possible, the memory system cannot simply be turned off, which means that memory will influence performance regardless of the presence of perceptual cues.

Our question here is whether activating an element of task knowledge on the participant's behalf moderates the relationship between cognitive ability and placekeeping performance. The task knowledge we activated is an acronym (a word in the English language) that represents a mnemonic for the correct step sequence. In the knowledge-activated condition, we took pains to make the acronym salient and to relate it to the correct sequence as participants learned the task. In the knowledge-not-activated condition, we took pains to disguise the acronym and to frustrate the discovery of any mnemonic that would help organize performance. Note that by "activated" and "not activated" we refer to the measures we took to either facilitate or frustrate early discovery of a relevant piece of task knowledge. Participants in the knowledge-not-activated condition may well have discovered (activated) their own knowledge for remembering the sequence, another point we return to in the *Discussion*.

We randomly assigned participants to either the knowledge-activated or knowledge-not-activated condition and then had them perform our placekeeping task, followed by a battery of cognitive ability tests. We then used regression analyses and structural equation modeling (SEM) to test the relationship between cognitive ability and placekeeping performance. The circumvention-of-limits hypothesis predicts that the relationship will be weaker in the knowledge-activated condition than in the knowledge-not-activated condition. To preview the results, we found that the relationship was in fact no weaker in the knowledge-activated condition than in the knowledge-not-activated condition.

EXPERIMENT

METHOD

Participants

The participants were undergraduate students recruited through the subject pool at Michigan State University. A total of 479 participants contributed data. At the data analysis stage, the following exclusions from this sample were made: We excluded 28 participants from the knowledge-activated condition and 27 participants from the knowledge-not-activated condition who had below-threshold accuracy (as further described in the *Procedure* section). Of the remaining 424 participants, we excluded 13 participants with outlying values, where an outlying score for a measure was defined as a score differing by more than 3.5 standard deviations from the total sample mean for that measure. We also excluded 1 participant with missing values and 83 participants who, as indicated by getting each of two relevant items wrong on a vocabulary test (described below), did not know the meaning of the word *unravel* and thus may not have used this word as a mnemonic in the knowledge-activated (*UNRAVEL*) condition (51% of these participants were nonnative English speakers). In total, 75 participants were excluded from the knowledge-activated condition and 77 participants were excluded from the knowledge-not-activated condition. Data from the remaining 327 participants were submitted to statistical analysis. The average self-reported ACT score (n = 295 reporting) was 25.9 (SD = 3.5), with a range from 17 to 34. The national average for this standardized college admissions test is approximately 21 (SD = 4), and thus the sample was selective in terms of cognitive ability.

Procedure

Participants performed all tasks in a single session lasting roughly 1 hr and 45 min. At the start of the session, participants reported their ACT scores and whether English was their first language. After this, participants performed, in order, the placekeeping task, Operation Span, Raven's Progressive Matrices, Number Series, Letter Sets, Perceptual Comparison Speed, and Vocabulary.[1] Between each test, the participant had a brief break as the experimenter started the next test. Participants were randomly assigned to complete the knowledge-activated or knowledge-not-activated version of the placekeeping task. We used a between-subject design for the knowledge activation manipulation because in a counterbalanced within-subject design the mnemonic would have transferred from the knowledge-activated condition when performed first to the knowledge-not-activated condition when performed second.

PLACEKEEPING.

In the placekeeping task (Altmann et al., 2014), participants performed 7 steps in a given sequence. Each step involved applying a two-alternative decision rule to a randomly generated stimulus. We refer to a performed step as a trial. After the last step of the sequence, the participant started the sequence over, generating a long sequence of trials. On every sixth trial, on average, performance was interrupted by a simple transcription typing task, after which the participant tried to resume the sequence at the correct step, namely the successor to the step performed immediately before the interruption.

Figure 1a includes two sample stimuli, each showing the features to which the decision rules were applied. The stimulus includes no information about which rule is currently correct—any of the 7 rules can be applied to any stimulus—so participants must remember where they are in the sequence. Figure 1b shows the rules themselves, in the form of screenshots of a help screen that was available to participants. The left side shows the help screen from the knowledge-activated condition, and the right side shows the help screen from the knowledge-not-activated condition; the only difference, which we discuss later, lies in the order of two terms in the fourth and fifth rules. To display the help screen, the participant pressed the "?" key (i.e., the Shift and "/" keys of a standard QWERTY keyboard) at any time during a trial. Pressing "?" caused the help screen to replace the regular stimulus, and releasing "?" caused the regular stimulus to replace the help screen. Figure 1c shows a sample stimulus from the interrupting task. There is a "code" that the participant must type, exactly, into a box. The code consists of the 14 different responses for the two-alternative decision rules (7 rules × 2 responses per rule), in random order. Participants typed two different codes per interruption. Any errors in typing a code were detected when the participant pressed the "Return" key after typing a code. An error caused the box to be cleared, after which the participant had to type that code again. The interrupting task took about 20 s.

On a given trial of the placekeeping task, participants responded to the stimulus by pressing a letter on the keyboard corresponding to the correct choice for the current decision rule. For example, if the current rule was "Underline-Italic" (Figure 1b), the participant pressed "u" if a character in the stimulus was underlined (as in the left stimulus in

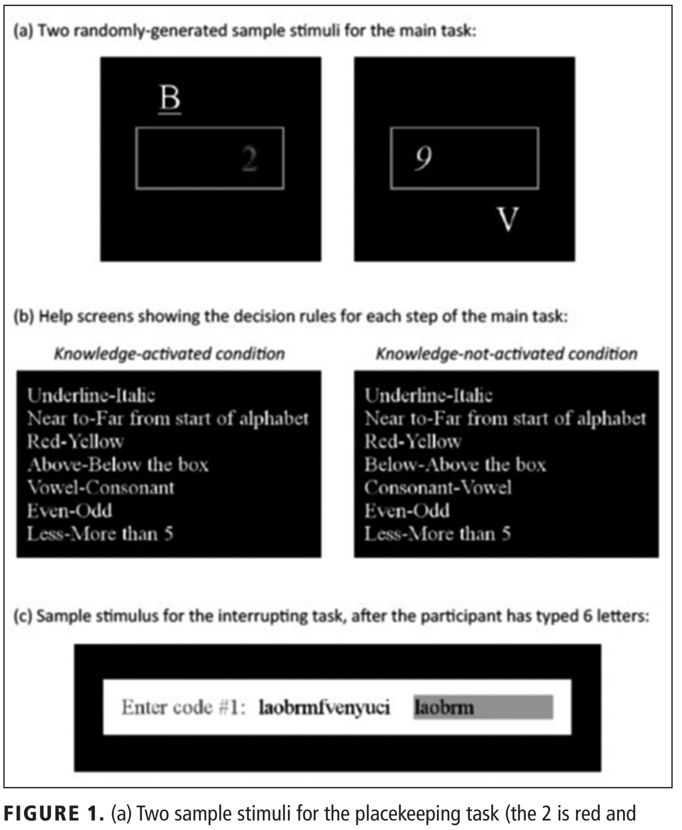

FIGURE 1. (a) Two sample stimuli for the placekeeping task (the 2 is red and the 9 is yellow). (b) Help screens showing the decision rules for the knowledge-activated condition (*UNRAVEL*; left) and knowledge-not-activated condition (*UNRBCEL*; right). (c) Sample stimulus for the interrupting task, after the participant has typed 6 letters of the first of two 14-letter "codes"

Figure 1a) or "i" if a character in the stimulus was in italics (as in the right stimulus in Figure 1a). On the next trial, the participant would apply the "Near to-Far from the start of the alphabet" rule, pressing "n" if the letter was near to the start of the alphabet (as in the left stimulus of Figure 1a) or "f" if the letter was far from the start of the alphabet (as in the right stimulus of Figure 1a). A trial took approximately 2.5 s to perform, with some variation between steps (see Altmann et al., 2014). Response time for trials included any time spent examining the help screen.

The placekeeping task began with a tutorial phase. During this phase, participants performed each of the seven steps once, in order, with instructions explaining the rule for that step, the two choices, and why the correct choice was correct. The tutorial phase was a locus of the knowledge activation manipulation, as we describe later. The tutorial phase was followed by a practice phase, during which the help screen in Figure 1b was visible at all times. At the end of the practice phase, the instructions indicated that the help screen would be hidden from then on but could be displayed at any time. At that point, the computer required the participant to display the help screen by pressing the "?" key, to ensure that the participant knew it was available and how to access it.

The practice phase was followed by the test phase, which contained 264 trials on average and exactly 40 interruptions. The test phase was organized into four blocks, each with 66 trials on average and exactly 10 interruptions. After each block, the participant was told their average response time (time per trial) and error rate for the block. The participant was also given feedback designed to encourage fast performance with accuracy above 70% but below ceiling (100%). The accuracy measure scored a trial as correct if the participant performed the correct step and selected the correct response for that step. A correct step was one that followed the step performed on the preceding trial in the rule sequence. A participant's data were excluded from analysis if their accuracy was below 70% on two or more blocks. The rationale for this criterion is that the participant was not following task instructions. If a participant's accuracy on a given block was below 70%, they received feedback asking them to be more accurate. Accordingly, accuracy below 70% for two or more blocks was evidence that the participant was not following instructions to be more accurate.

The knowledge activation manipulation was implemented in two ways: in terms of the help screen, as shown in Figure 1b, and in terms of the tutorial phase at the start of the placekeeping task. In the knowledge-activated condition, the instructions at the start of the tutorial phase drew attention to the mnemonic and its representation of the correct step sequence. Specifically, the computer displayed the following passage:

> Your task is to apply one of a set of simple rules to each stimulus. There are seven different rules, and you apply them in a particular order. To remember the order, you have to remember one word: UNRAVEL. Each letter of UNRAVEL is the first letter of one of the rules. In the next few minutes you'll get to see how this works.

As participants performed each step during the tutorial phase, the instructions for that step drew their attention to the role of the name for that step (U, N, R, A, V, E, or L) in extending the acronym by one letter.

In the knowledge-not-activated condition, we took measures to frustrate the discovery of any mnemonic for the correct order of the steps. One such measure was to disguise the UNRAVEL acronym by altering the format of two of the decision rules. The knowledge-activated variants of these rules were "Above-Below" and "Vowel-Consonant," whereas the knowledge-not-activated variants were "Below-Above" and "Consonant-Vowel." The effect of reversing the terms within each rule can be seen in the two variants of the help screen in Figure 1b. In the knowledge-activated variant, on the left, the letters down the left margin spell the word *UNRAVEL,* whereas in the knowledge-not-activated variant, on the right, the letters down the left margin spell the not-easily-pronounceable nonword *UNRBCEL*. Thus, the help screen reinforced the acronym in one condition but not in the other.

In this condition, we also omitted any mention of an acronym (actual or hypothetical) during the tutorial phase. Thus, the passage corresponding to the one above read simply,

> Your task is to apply one of a set of simple rules to each stimulus. There are seven different rules, and you apply them in a particular order. In the next few minutes you'll get to see how this works.

As the participant performed each step during the tutorial phase, the instructions said nothing to hint at the possibility of forming an acronym from the letters of the responses, and the "Below-Above" and "Consonant-Vowel" variants of those two rules were used, helping to disguise the *UNRAVEL* acronym. With these measures in place, participants in the knowledge-not-activated condition were left to their own devices to implement the step sequence correctly. At the same time, the step sequence was exactly matched to that in the knowledge-activated condition, and, across conditions, most (five of seven) of the decision rules were presented in the identical format. We expected that use of the help screen would be more frequent in the condition without the mnemonic, and we measured frequency of help use as a manipulation check.

For each participant in the two conditions, we computed the average error rate and the average response time across the four blocks. Coefficient α and other descriptive statistics for this task are discussed in the *Results* section.

OPERATION SPAN.

We used the automated version of the operation span task to measure working memory capacity (Unsworth, Heitz, Schrock, & Engle, 2005). The participant's task is to solve a series of math equations and, after solving each equation, to remember a letter. After three to seven equation–letter trials, participants are prompted to recall the letters in the order in which they were presented. There were 15 sets, with 3 sets at each length. The measure of working memory capacity was the number of letters recalled in the correct order (maximum = 75). The coefficient α for operation span was .78.

RAVEN'S PROGRESSIVE MATRICES.

This test of fluid intelligence consists of the 18 odd-numbered items from Raven's Advanced Progressive Matrices (Raven, Raven, & Court, 1998). Each item is displayed on a computer screen and consists of a series of patterns arranged in three rows and three columns. The pattern in the lower-right is always missing, and the participant's task is to choose the alternative that logically completes the series. The time limit is 10 min, and the measure is the number correct. The coefficient α for Raven's was .65.

NUMBER SERIES.

This test of fluid intelligence consists of 15 items from the test of primary mental abilities (Thurstone, 1938). Each item is displayed on a computer screen and consists of a series of numbers; the participant's task is to choose which of four alternatives logically completes the series. The time limit is 4.5 min, and the measure is the number correct. The coefficient α for Number Series was .70.

LETTER SETS.

This test of fluid intelligence consists of 20 items from the letter sets test of the Educational Testing Service Kit of Factor-Referenced Cognitive Tests (Ekstrom, French, Harman, & Dermen, 1976). Each item is displayed on a computer screen and consists of five sets of four letters (e.g., *BCDE*) arranged in a row. The participant's task is to choose the set that is unlike the other four sets in terms of the rule governing the relations between the letters. The time limit was 10 min, and the measure was the number correct. The coefficient α for Letter Sets was .69.

LETTER AND NUMBER COMPARISON.

There were two versions of this test of perceptual speed. In Letter Comparison (Salthouse & Babcock, 1991), the participant's task was to judge whether sets of three, six, or nine consonants on either side of a line were the same or different. If the sets were the same, participants clicked "SAME"; if the sets differed, participants clicked "DIFFERENT." Participants had 30 s to complete as many comparisons as possible. In Number Comparison, the task was the same, except that the stimuli were sets of three, six, or nine digits. In each test, there were two 30-s administrations, and the measure was the average number correct minus the number incorrect across the two administrations. The coefficient α was .75 for Letter Comparison and .75 for Number Comparison, as computed using the two forms of each test.

VOCABULARY.

This test consists of 20 multiple-choice items to assess word knowledge, including 10 synonym items and 10 antonym items (Hambrick, Salthouse, & Meinz, 1999). We added an eleventh synonym item and an eleventh antonym item that assessed the meaning of the word *unravel*. Each item consists of a target word and four words that serve as alternatives. For synonym items, the participant's task is to choose the alternative most nearly the same in meaning to the target word, whereas for antonym items, it is to choose the alternative most nearly the opposite in meaning to the target word. The score for each section (synonym and antonym) was the number correct. The coefficient α was .50 for Synonym and .42 for Antonym.

Cognitive Ability Factors

We entered the cognitive ability measures into an exploratory factor analysis (principal axis extraction). As expected, the first unrotated factor accounted for a sizeable amount of the variance (29.6%); we saved

TABLE 1. Descriptive Statistics and Factor Loadings for Cognitive Ability Measures

				Factor		
	M	SD	α	I	II	III
Operation Span	60.0	10.6	.78	**.53**	–.05	–.11
Raven's Matrices	9.5	2.9	.65	**.64**	–.02	.06
Number Series	9.0	2.4	.70	**.42**	.20	.01
Letter Sets	10.8	2.6	.69	**.64**	–.05	.07
Letter Comparison	20.0	4.2	.75	–.08	**.69**	.05
Number Comparison	30.7	4.7	.75	.06	**.88**	–.05
Synonyms	3.2	2.0	.50	–.10	.03	**.83**
Antonyms	3.6	1.8	.42	.15	–.04	**.50**

Note. *N = 327. Factors correlations (rs): I–II = .20; I–III = .44; II–III = .05. Rotation procedure: Promax. Factor extraction: principal axis. Factor loadings with an absolute value greater than .30 are in bold.*

scores for this factor as the estimate of the general factor of intelligence (*g*) for the regression analyses reported in the following section. We then repeated this analysis to determine how many cognitive ability factors were present in the data, with an oblique rotation (Promax), allowing for interfactor correlations (see Fabrigar, Wegener, MacCallum, & Strahan, 1999). Three factors met the criterion for extraction (eigenvalue > 1); these factors were clearly interpretable as reflecting fluid intelligence, perceptual speed, and crystallized intelligence; as we did for *g*, we saved factor scores for use as estimates of these cognitive abilities in regression analyses. See Table 1 for descriptive statistics and factor loadings; means for the cognitive ability measures also did not differ across conditions (all ps > .05).

RESULTS

Descriptive statistics for placekeeping performance are displayed in Table 2. Consistent with our previous studies (Altmann et al., 2014; Hambrick & Altmann, 2015), there was a considerable amount of variability in error rate: knowledge-not-activated (0% to 16%; M = 5.7%, SD = 3.5%) and knowledge-activated (0% to 19%; M = 5.3%, SD = 3.8%). This was true for response time also: knowledge-not-activated (2.0 to 5.1 s; M = 3.1 s, SD = 0.5 s) and knowledge-activated (1.8 to 5.5 s; M = 3.2 s, SD = 0.6 s). For both error rate and response time, skewness and kurtosis statistics were in the acceptable range (<2) for both conditions, and therefore we used untransformed rather than transformed variables for all subsequent analyses. Coefficient α, using the four blocks as the variables, ranged from .63 to .84 (see Table 2), indicating generally acceptable internal consistency reliability.

Participants used the help screen significantly more often in the knowledge-not-activated condition (48% of trials) than in the knowledge-activated condition (33% of trials), t = 4.36, d = 0.48, p < .001. Moreover, as shown in Figure 2, the distribution of help use was strikingly different across conditions.

TABLE 2. Descriptive Statistics for Placekeeping Performance

	Knowledge-not-activated			Knowledge-activated			
	M	*SD*	*α*	*M*	*SD*	*α*	*t*
Error rate (%)	5.7	3.5	.63	5.3	3.8	.73	1.10
Response time (s)	3.1	0.5	.84	3.2	0.6	.84	–0.88

Note. *Knowledge-not-activated (UNRBCEL)* n *= 163; knowledge-activated (UNRAVEL)* n *= 164. For reliability analyses,* n *= 155 and* n *= 154, respectively (some participants were missing values at the block level).*

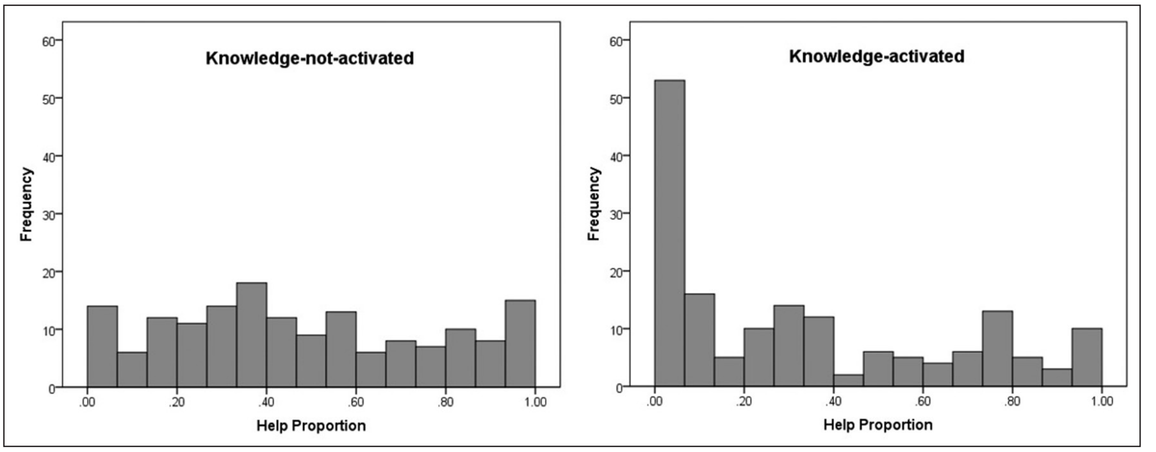

FIGURE 2. Frequency distributions for proportion of trials on which help was used for the knowledge-not-activated (*UNRBCEL*) condition (left panel; $n = 163$) and knowledge-activated (*UNRAVEL*) condition (right panel; $n = 164$)

In the knowledge-activated condition, more than a quarter of participants (28%) used the Help function on no more than 5% of trials, indicating that they relied largely on the mnemonic to perform the task. By contrast, in the knowledge-not-activated condition, only 6.7% of participants used the Help function on no more than 5% of trials, and the distribution of help use was essentially uniform across the full range of the variable, indicating that different strategies were used. A nonparametric test indicated that the distributions differed across conditions (Mann–Whitney $U = 9{,}251$, two-tailed $p < .001$). We take these effects on frequency of help use as evidence that the knowledge activation manipulation was successful.

We had also expected that the knowledge activation manipulation would have an effect on accuracy, with more accurate performance in the knowledge-activated condition. There was no difference in placekeeping performance across the two conditions, in error rate or response time. We examine the implications of these null effects in the *Discussion*.

Correlations between the major variables are presented in Table 3. The correlations between the cognitive ability and placekeeping measures were generally negative, indicating better performance (lower error rates, faster responding) for high-ability participants. The correlations also did not generally differ across conditions, with only one exception.

TABLE 3. Correlations between Placekeeping Performance Measures and Cognitive Ability Factors

	Error rate			Response time		
Ability Factor	Knowledge-not-activated	Knowledge-activated	*z*	Knowledge-not-activated	Knowledge-activated	*z*
g	–.10	–.25**	1.39	–.39**	–.39*	0.00
Fluid intelligence	–.20*	– .28**	0.76	–.33**	–.31*	–0.20
Perceptual speed	.10	–.13	2.07*	–.32**	–.42**	1.04
Crystallized intelligence	–.05	–.08	0.27	–.19*	–.12	–0.64

Note. *Knowledge-not-activated (*UNRBCEL*)* n *= 163. Knowledge-activated (*UNRAVEL*)* n *= 164.* g*, general intelligence,* z*, test of difference between correlations across conditions.*
*p *< .05,* **p *< .01.*

Regression Analyses

We performed a hierarchical regression analysis to test whether the relationship between general intelligence (*g*) and placekeeping performance was weaker in the knowledge-activated condition than in the knowledge-not-activated condition, as predicted by the circumvention-of-limits hypothesis. With separate analyses for average error rate and average response time, we regressed placekeeping performance onto *g* and a dummy-coded variable for condition (Step 1), and then onto a condition × *g* interaction term, created by multiplying the condition variable and *g* (Step 2). We then repeated this analysis, using as predictor variables fluid intelligence, perceptual speed, and crystallized intelligence (Step 1) and condition × ability interaction terms (Step 2).

Results are displayed in Tables 4 and 5. There was a main effect of *g* (Table 4) on both error rate ($\beta = -.18$, $t = -3.22$, $p < .01$) and response time ($\beta = -.39$, $t = -7.58$, $p < .01$). There was also a main effect of fluid intelligence (Table 5) on error rate ($\beta = -.31$, $t = -4.65$, $p < .01$), and main effects of fluid intelligence ($\beta = -.24$, $t = -3.75$, $p < .01$) and perceptual speed on response time ($\beta = -.31$, $t = -5.88$, $p < .01$). There was also a significant condition × perceptual speed interaction on error rate ($\beta = -.11$, $t = -2.04$, $p < .05$), but the interaction was in the unexpected direction of perceptual speed having a positive effect in the knowledge-not-activated condition but a negative effect in the knowledge-activated condition (see correlations in Table 3). No other interactions were statistically significant. In general, high levels of cognitive ability were associated with superior placekeeping performance across conditions.

Structural Equation Modeling

We used multiple-groups SEM to further test for differential effects of cognitive ability on placekeeping performance across the knowledge-not-activated and knowledge-activated conditions. With the covariance matrix as the input for the analysis, SEM is a tool for evaluating models stipulating relations between variables. We characterize model fits using a number of commonly-reported fit statistics (see Kline, 2015), including the chi-square (χ^2) statistic, the comparative fit index (CFI) and normed fit index (NFI), and the root mean square error of approximation (RMSEA). The χ^2 test assesses deviation of

TABLE 4. Results of Hierarchical Regression Analysis with *g* Factor Predicting Placekeeping Performance

	ΔR^2	ΔF	β	t	sr^2
DV = Error rate					
Step 1	.035	5.95**			
Condition			–.07	–1.27	.005
g			–.18	–3.22**	.031
Step 2	.006	1.89			
Condition × *g*			–.08	–1.38	.006
DV = Response time					
Step 1	.153	29.35**			
Condition			.03	0.58	.001
g			–.39	–7.58**	.151
Step 2	.000	0.01			
Condition × *g*			–.01	–0.11	.000

Note. *Condition dummy-coded as knowledge-not-activated (*UNRBCEL*) = –1, knowledge-activated (*UNRAVEL*) = 1. Step 1* df *= 2,324; Step 2* df *= 1,323.* β *= standardized regression coefficient; DV = dependent variable;* g *= general intelligence;* sr^2 *= squared semipartial correlation, reflecting independent contribution of the predictor to the outcome.*
*p < .05. **p < .01.

TABLE 5. Results of Hierarchical Regression Analysis With Broad Cognitive Abilities Predicting Placekeeping Performance

	ΔR^2	ΔF	β	t	sr^2
DV = Error rate					
Step 1	.070	6.08**			
Condition			–.06	–1.16	.004
Fluid Intelligence			–.31	–4.65**	.062
Perceptual Speed			.06	1.03	.003
Crystallized Intelligence			.11	1.62	.008
Step 2	.014	1.63			
Condition × fluid intelligence			–.01	–0.20	.000
Condition × perceptual speed			–.11	–2.04*	.012
Condition × crystallized intelligence			–.01	–0.80	.000
DV = Response time					
Step 1	.193	19.25**			
Condition			.03	0.68	.001
Fluid Intelligence			–.24	–3.75**	.035
Perceptual Speed			–.31	–5.88**	.087
Crystallized Intelligence			–.01	–0.10	.000
Step 2	.004	0.55			
Condition × fluid intelligence			–.01	–0.16	.000
Condition × perceptual speed			–.05	–0.92	.002
Condition × crystallized intelligence			.05	0.78	.002

Note. *Condition dummy-coded as knowledge-not-activated (*UNRBCEL*) = –1, knowledge-activated (*UNRAVEL*) = 1. Step 1* df *= 4,322; Step 2* df *= 3,319. β = standardized regression coefficient;* sr^2 *= squared semipartial correlation, reflecting independent contribution of the predictor to the placekeeping measure.*
*$p < .05$. **$p < .01$.

the covariance matrix implied by the hypothesized model and the covariance matrix generated from the data. Nonsignificant values are therefore desirable, but even slight deviations can result in significant values. The CFI and NFI, which are less sensitive to sample size, are "baseline" fit statistics and reflect improvement of model fit over a model in which the covariances are assumed to be zero. CFI and NFI values greater than .90 indicate "acceptable" fit. Finally, the RMSEA reflects the difference between the model-implied and observed covariance matrices, corrected for degrees of freedom. Values less than .08 are considered acceptable.

Three major steps were involved in this analysis. With condition as the grouping variable, the first step was to specify a measurement model for the predictor variables and one for the outcome variables. In the predictor model, the latent variables were fluid intelligence (indicators: Letter Sets, Number Series, Raven's, and Operation Span), crystallized intelligence (indicators: Antonyms and Synonyms), and perceptual speed (indicators: Blocks A and B of Letter Comparison, Blocks A and B of Number Comparison). In the outcome model, the latent variables were error rate (indicators: Blocks 1–4) and response time (indicators: Blocks 1–4).

The second step was to test for different forms of measurement equivalence across the two conditions. In other words, we tested whether the same measurement model provided a good fit to the data in the two conditions, permitting an "apples to apples" comparison across conditions. We tested for configural equivalence by examining the fit of the models when the number of factors and the factor indicator correspondences were constrained across conditions, but the factor loadings were unconstrained (Kline, 2015). Overall, the fit of the predictor model was acceptable, $\chi^2(64) = 111.37$, $p < .001$, CFI = .93, NFI = .86, RMSEA = .05, as was the fit of the outcome model, $\chi^2(38) = 115.41$, $p < .001$, CFI = .92, NFI = .89, RMSEA = .08. Next, we tested for metric equivalence by assessing the change in model fit after constraining factor loadings to be equal across conditions. The χ^2 difference test was not significant for the predictor model, $\chi^2(7) = 6.22$, $p = .51$, or for the outcome model, $\chi^2(6) = 4.23$, $p = .65$. Thus, the analyses establish configural and metric equivalence for both the predictor and outcome models. This indicates that the latent factors represented in the models are comparable across conditions.

With this established, the final step of the SEM was to test for differences in relationships between predictor and outcome variables across conditions by constraining unidirectional paths between latent variables to be equivalent. If constraining these paths to be equivalent across conditions results in a significant loss of model fit, this would indicate a significant difference in the paths across conditions. In this analysis, we first estimated the model with no constraints, and then constrained individual paths from predictor to outcome variables one at a time. The final model is shown in Figure 3 (if a pair of

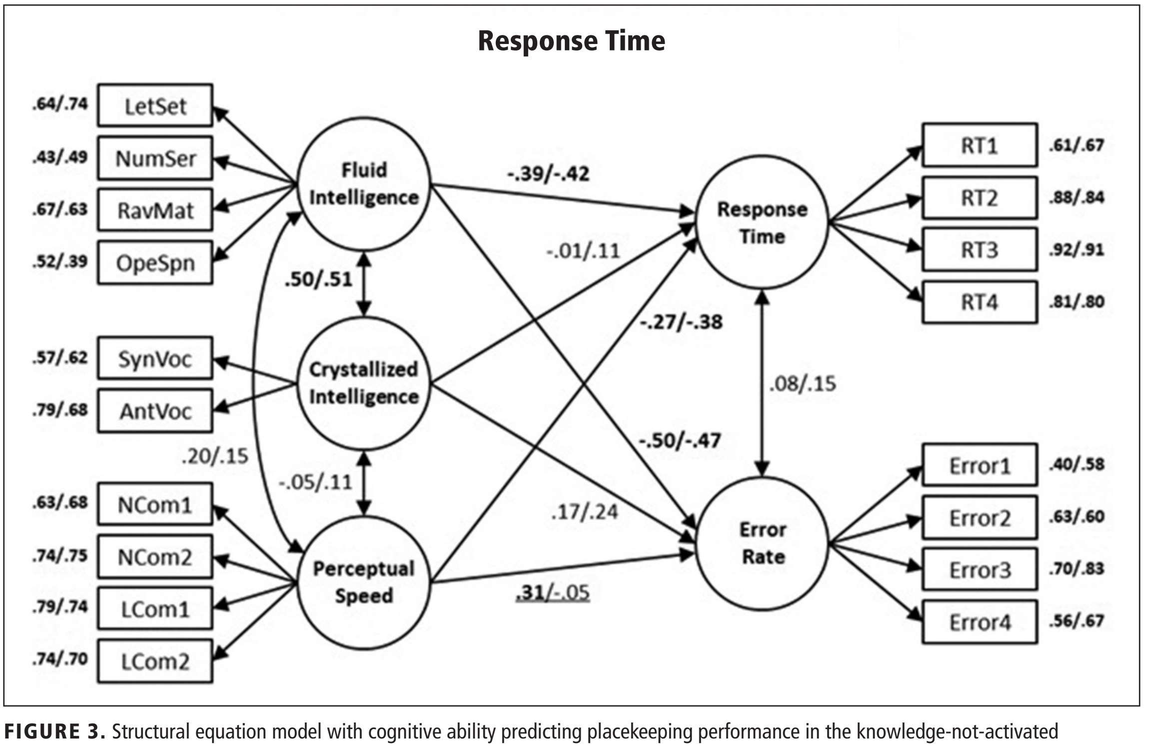

FIGURE 3. Structural equation model with cognitive ability predicting placekeeping performance in the knowledge-not-activated (*UNRBCEL*) condition ($n = 163$) and knowledge-activated (*UNRAVEL*) condition ($n = 164$). AntVoc = Antonym Vocabulary; LCom = Letter Comparison; LetSet = Letter Sets; NCom = Number Comparison; NumSer = Number Series; OpeSpn = Operation Span; RavMat = Raven's Matrices; RT = response time; SynVoc = Synonym Vocabulary. Rectangles represent observed variables; circles represent latent variables. Values adjacent to the observed variables are factor loadings. Values adjacent to unidirectional arrows are standardized path coefficients; values adjacent to bidirectional arrows are correlations. Values in bold are statistically significant ($p < .05$). For each pair of values, the left value is for the knowledge-not-activated condition (*UNRBCEL*), and the right value is for the knowledge-activated condition (*UNRAVEL*); the pair of values underlined are significantly different from each other

values are underlined, that indicates they are significantly different from each other, $p < .05$). Consistent with the preceding results, Fluid Intelligence had a significant negative relationship with both Error Rate and Response Time, in both conditions: knowledge-not-activated (–.50, $p < .05$, and –.39, $p < .01$, respectively) and knowledge-activated (–.47 and –.42, ps < .01, respectively). The relationship between Perceptual Speed and Error Rate was significantly larger in the knowledge-not-activated condition (.31, $p < .05$) than in the knowledge-activated condition (–.05, *ns*), $\Delta\chi^2(1) = 5.34$, $p = .02$, but this difference is not consistent with the circumvention-of-limits hypothesis, as we discuss later. There were no other significant differences. Overall, the fit of the model was acceptable, $\chi^2(250) = 392.46$, $p < .001$, CFI = .92, NFI = .82, RMSEA = .04.

DISCUSSION

The circumvention-of-limits hypothesis implies that performance limitations associated with general cognitive ability can be overcome through training. Here, we simulated effects of training by experimentally manipulating whether or not a piece of task-relevant knowledge—namely, a word in the English language—was activated at the start of the experimental session. In the knowledge-activated condition, we gave participants a mnemonic for the correct order of steps in the task sequence. In the knowledge-not-activated condition, we did not give them a mnemonic and frustrated the discovery of any mnemonic. The knowledge activation manipulation had a significant effect on the frequency of help use, which was higher in the knowledge-not-activated condition. We take this effect as evidence that the manipulation had the intended effect. Furthermore, although some participants in the knowledge-not-activated condition may have developed their own mnemonics or other memory strategies to remember the steps, the distribution of help use in this condition (see Figure 2) indicates quite clearly that most participants did *not* do so. For example, 44% of participants in this condition used help on at least half of the trials, and 24% used help on at least 75% of the trials. The knowledge activation manipulation had no effect on performance as measured by accuracy and response time, which is an interesting outcome that we discuss later.

Of primary interest was whether the knowledge activation manipulation moderated the relationship between general cognitive ability and task performance. According to the circumvention-of-limits hypothesis, the relationship between general cognitive ability and task performance should have been weaker in the knowledge-activated condition than in the knowledge-not-activated condition, reflecting the hypothetical role of additional knowledge (the activated mnemonic) compensating for ability. However, at the level of *g* and at the level of the broad ability factors of fluid intelligence and crystallized intelligence, the relationship did not differ across conditions. That is, high ability was as predictive of performance in the knowledge-activated condition as in the knowledge-not-activated condition. Accordingly, the circumvention-of-limits hypothesis was not supported.

We did find a significant interaction between condition and perceptual speed, but this interaction reflected a positive relationship between perceptual speed and error rate in the knowledge-not-activated condition. That is, in this condition people high in this factor actually performed *worse* than people low in this factor (faster processing was associated with higher error rate). Although this relationship would be intriguing if it replicated, we can think of no obvious explanation for it and do not attempt to interpret it.

An interesting outcome was that even though the knowledge activation manipulation affected frequency of help use (Figure 2), it had no effect on error rate or response time. This pattern of results suggests that knowledge-not-activated participants found ways to compensate for the lack of the mnemonic that knowledge-activated participants had available, in part by using the help screen more often but also potentially by devising their own strategies to memorize the correct sequence of steps. If knowledge-not-activated participants developed compensatory strategies like this, then the results of this study are inconsistent with the strategy mediation hypothesis (Bailey, Dunlosky, & Kane, 2011). According to this hypothesis, cognitive ability measures predict complex task performance not because they capture basic capacities (or "primitives") but rather because they measure a person's skill in devising strategies to perform cognitive tasks (e.g., McNamara & Scott, 2001). Accordingly, this hypothesis leads to the prediction that experimental manipulations that minimize individual differences in strategy use in a cognitive ability task should decrease the correlation between performance in that task and outcomes. Our knowledge activation manipulation might be considered such a manipulation, and to reiterate, it had essentially no effect on the relationship between placekeeping performance and the other cognitive measures.

The main limitations on our results are limitations associated with the knowledge activation approach generally. One limitation is that the knowledge gap created by activating a simple and common memory representation in one group of participants but not the other is almost certainly smaller than the knowledge gap between highly trained experts and novices in a complex task domain. For example, we activated a mnemonic for placekeeping, but placekeeping is only one cognitive operation in our task. Another important operation, for example, is remembering the step performed before an interruption, which relies on episodic memory for recent performance rather than memory for the step sequence. In more complex task domains, the knowledge gap will be bigger than in our simpler task. For example, a skilled pilot may use placekeeping to remember the sequence of steps for takeoff, but the steps themselves (e.g., manipulating controls, communicating with air traffic control) are themselves likely to be performed better by experts than by novices.

Another limitation involves individual differences in the structure and content of expert knowledge, which the knowledge activation approach is not designed to capture. One example involves retrieval structures, which allow storage and retrieval of intermediate products in a complex task (Ericsson & Kintsch, 1995). Our participants may have developed retrieval structures for remembering what step of the task sequence they performed last before an interruption, so they could know where to resume after the interruption. However, the development and use of such structures are not factors we could manipulate or measure. In expert performance, retrieval structures can be highly idiosyncratic as a function of individuals' preexisting knowledge. These structures, and other similarly individualized representations and processes developed through training, may be means by which experts compensate for limits on general cognitive ability but that are difficult to manipulate experimentally when all members of a group of participants have to be treated the same way.

These limitations aside, it remains the case that the knowledge activation approach affords experimental control through random assignment. The approach thus addresses a limitation that is inherent in bona fide expert–novice comparisons, which is that they are quasiexperimental and thus subject to a variety of confounding factors (e.g., age, as we noted earlier). Thus, we argue that converging operations using multiple methods, including knowledge activation, play an important role in developing evidence on the relationship between knowledge, ability, and performance.

Beyond the present study and the one by Hambrick and Oswald (2005), we know of one other study that has used the knowledge activation approach. Bermingham, Gardner, and Woltz (2016) had participants complete tests of working memory capacity and then perform an episodic memory test that involved remembering either fictitious food items and their prices (e.g., "Bourf, $2.22") or actual food items and their prices. The latter condition allowed participants to draw on their knowledge of grocery items and their prices. Working memory capacity had a weaker relationship with recall of the actual items than with recall of the fictitious items, supporting the circumvention-of-limits hypothesis. However, the predictor and criterion tasks may have been more similar to each other in the knowledge-not-activated condition than in the knowledge-activated condition, because in the knowledge-not-activated condition they both involved remembering random strings of digits. This greater similarity of the tasks may have accounted for the greater correlation in performance across the two tasks in this condition.

The dominant view, dating at least to Chase and Simon (1973), has been that expert performance largely reflects training history rather than ability, but in fact the evidence for the circumvention-of-limits hypothesis is mixed (see Hambrick, Burgoyne, & Oswald, in press, for a review). Some studies have supported the hypothesis. Sohn and Doane (2003) found that a measure of long-term memory skill reduced the relationship between working memory capacity and performance of a situational awareness task in pilots. Similarly, Hambrick et al. (2012) found that geological knowledge attenuated the relationship between visuospatial ability and performance in a geological bedrock-mapping task. Other studies have not supported the hypothesis. Meinz and Hambrick (2010) found that working memory capacity had as strong a relationship with sight-reading performance in pianists at low levels of deliberate practice as at higher levels. Similarly, Morrow, Menard, Stine-Morrow, Teller, and Bryant (2001) found no evidence that aviation expertise mitigated effects of cognitive ability on pilots' performance in aviation-related communication tasks. Moreover, as we noted earlier, interpretation of evidence from quasiexperimental tests of the circumvention-of-limits hypothesis is complicated by the fact that such studies are vulnerable to confounds.

The circumvention-of-limits hypothesis is an appealing idea (see Ericsson & Pool, 2016) with far-reaching implications for the question of what people can achieve through training. Our results add to the evidence calling this view into question and suggesting that cognitive ability may, in fact, impose limits on performance that cannot always be overcome through training. One important direction for future work is to develop a theoretical framework for making testable predictions about when cognitive ability factors should affect performance in different tasks and when they should not. This theoretical framework could be based on computational models of task environments (Gobet, 2016), task analyses (Chipman, Schraagen, & Shalin, 2000), or both and could be empirically tested using both quasiexperimental and experimental research designs. It is entirely possible that circumvention of limits associated with cognitive ability is possible in some types of cognitive tasks. Another important direction for future work is to develop more powerful knowledge activation manipulations, perhaps extending this approach to study processes such as pattern recognition that are believed to be important in a wide range of sports and games. We believe that this sort of programmatic approach to research on the interplay between domain-general and domain-specific factors is a critical next step in research on the underpinnings of human expertise.

NOTES

This research was funded by grants from the Office of Naval Research (N00014-13-1-0247, N00014-16-1-2457, and N00014-16-1-2841). We thank Laura Bell and Kylie Smith for help with data collection.

Address correspondence about this article to David Z. Hambrick or Erik M. Altmann, Department of Psychology, Michigan State University, East Lansing, MI 48824 (e-mail: hambric3@msu.edu, ema@msu.edu).

1. Participants also completed a personality inventory, which is beyond the scope of this article.

REFERENCES

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.

Bailey, H., Dunlosky, J., & Kane, M. J. (2011). Contribution of strategy use to performance on complex and simple span tasks. *Memory & Cognition, 39,* 447–461.

Bermingham, D., Gardner, M. K., & Woltz, D. J. (2016). Effects of working memory capacity and domain knowledge on recall for grocery prices. *American Journal of Psychology, 129,* 149–160.

Burgoyne, A. P., Sala, G., Gobet, F., Macnamara, B. N., Campitelli, G., & Hambrick, D. Z. (2016). The relationship between cognitive ability and chess skill: A comprehensive meta-analysis. *Intelligence, 59,* 72–83.

Chase, W. G., & Simon, H. A. (1973). Perception in chess. *Cognitive Psychology, 4,* 55–81.

Chipman, S. F., Schraagen, J. M., & Shalin, V. L. (2000). Introduction to cognitive task analysis. In J. M. Schraagen, S. F. Chipman, & V. L. Shalin (Eds.), *Cognitive task analysis* (pp. 3–23). New York, NY: Erlbaum.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Kit of Factor-Referenced Cognitive Tests*. Princeton, NJ: Educational Testing Service.

Ericsson, K. A. (2014). Why expert performance is special and cannot be extrapolated from studies of performance in the general population: A response to criticisms. *Intelligence, 45,* 81–103.

Ericsson, K. A., & Charness, N. (1994). Expert performance: Its structure and acquisition. *American Psychologist, 49,* 725–747.

Ericsson, K. A., & Kintsch, W. (1995). Long-term working memory. *Psychological Review, 102,* 211–245.

Ericsson, A., & Pool, R. (2016). *Peak: Secrets from the new science of expertise*. New York, NY: Houghton Mifflin Harcourt.

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods, 4,* 272–299.

Fitts, P. M., & Posner, M. I. (1967). *Human performance.* Oxford, England: Brooks/Cole.

Gobet, F. (2016). *Understanding expertise: A multi-disciplinary approach.* Basingstoke, England: Palgrave.

Gobet, F., & Campitelli, G. (2007). The role of domain-specific practice, handedness, and starting age in chess. *Developmental Psychology, 43,* 159–172.

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

Hambrick, D. Z., Burgoyne, A. P., Oswald, F. L. (in press). Domain-general models of expertise: The role of cognitive ability. In P. Ward, J. M. Schraagen, J. Gore, & E. Roth (Eds.), *Oxford handbook of expertise: Research and application*. Oxford, England: Oxford University Press.

Hambrick, D. Z., Libarkin, J. C., Petcovic, H. L., Baker, K. M., Elkins, J., Callahan, C. N., . . . LaDue, N. D. (2012). A test of the circumvention-of-limits hypothesis in scientific problem solving: The case of geological bedrock mapping. *Journal of Experimental Psychology: General, 141,* 397–403.

Hambrick, D. Z., & Meinz, E. J. (2011). Limits on the predictive power of domain-specific experience and knowledge in skilled performance. *Current Directions in Psychological Science, 20,* 275–279.

Hambrick, D. Z., & Oswald, F. L. (2005). Does domain knowledge moderate involvement of working memory capacity in higher-level cognition? A test of three models. *Journal of Memory and Language, 52,* 377–397.

Hambrick, D. Z., Oswald, F. L., Altmann, E. M., Meinz, E. J., Gobet, F., & Campitelli, G. (2014). Deliberate practice: Is that all it takes to become an expert? *Intelligence, 45,* 34–45.

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.

Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). New York, NY: Guilford.

McNamara, D. S., & Scott, J. L. (2001). Working memory capacity and strategy use. *Memory &Cognition, 29,* 10–17.

Meinz, E. J., & Hambrick, D. Z. (2010). Deliberate practice is necessary but not sufficient to explain individual differences in piano sight-reading skill the role of working memory capacity. *Psychological Science, 21,* 914–919.

Morrow, D. G., Menard, W. E., Stine-Morrow, E. A. L., Teller, T., & Bryant, D. (2001). The influence of task factors and expertise on age differences in pilot communication. *Psychology and Aging, 16,* 31–46.

Raven, J., Raven, J. C., & Court, J. H. (1998). *Manual for Raven's progressive matrices and vocabulary scales: Section 4. The advanced progressive matrices*. San Antonio, TX: Harcourt Assessment.

Salthouse, T. A. (1991). Expertise as the circumvention of human processing limitations. In K. A. Ericsson & J. Smith (Eds.), *Toward a general theory of expertise: Prospects and limits* (p. 286). Cambridge, UK: Cambridge University Press.

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

Sohn, Y. W., & Doane, S. M. (2003). Roles of working memory capacity and long-term working memory skill in complex task performance. *Memory & Cognition, 31,* 458–466.

Thurstone, L. L. (1938). *Primary mental abilities*. Chicago, IL: University of Chicago Press.

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