Astronaut Candidate, Candidate-Like, and Undergraduate Subjects Compared on Retention and Transfer

James A. Kole; Immanuel Barshi; Alice F. Healy; Vivian I. Schneider

INTRODUCTION: The present study examined long-term retention and transfer of knowledge and skills, as well as the effect of cognitive load on retention and transfer, using a sample of astronaut candidates and two comparison groups. The first comparison group, recruited from Johnson Space Center, was similar in age, education, and general health to the astronaut candidate group; the second comparison group included university undergraduate students.

METHODS: This study employed two different tasks—a simple perceptual-motor task involving data entry and a complex memory updating task requiring both prospective and retrospective memory. Subjects completed multiple sessions involving both tasks over a 500-d period, with test sessions involving transfer and/or a cognitive load manipulation. For the perceptual-motor task, transfer involved changes to the stimuli that increased intrinsic cognitive load or changes to the required motoric procedures. For the memory updating task, extraneous cognitive load was increased by the addition of a concurrent secondary task.

RESULTS: For both the perceptual-motor and memory updating tasks, astronaut candidates and candidate-like subjects performed more accurately, with greater speed, and were less impacted by increased cognitive load than undergraduate students. Despite the generally superior performance of astronaut candidates and candidate-like subjects, they were more likely to experience negative transfer on the perceptual-motor task, whereas undergraduate students demonstrated positive transfer.

DISCUSSION: Candidate-like subjects provided a more accurate approximation of astronaut candidate performance than did undergraduate students, especially with regard to negative transfer effects and cognitive load.

KEYWORDS: retention, transfer of learning, astronaut.

Kole JA, Barshi I, Healy AF, Schneider VI. *Astronaut candidate, candidate-like, and undergraduate subjects compared on retention and transfer*. Aerosp Med Hum Perform. 2023; 94(12):902–910.

The space environment imposes significant physiological, psycho-social, and cognitive loads on astronaut crew that likely impact crew performance during missions. To date, no systematic data collection has taken place to understand the effects of such loads on astronauts' ability to retain trained knowledge and skills, and to transfer such knowledge and skills to novel situations. The present study is the first to systematically collect data on the effects of cognitive loads on long duration retention and transfer.

To become an astronaut, several educational and physical requirements must be met. Astronaut candidates must hold at minimum a master's degree in a science, technology, engineering, or math (STEM) field, or a bachelor's degree in a STEM field with completion of a test pilot program. Physical requirements focus on basic indices of health, such as blood pressure and visual acuity, as well as height and weight restrictions due to spacesuit and spacecraft constraints. In 2016 there were

From the University of Northern Colorado, Greeley, CO, United States; NASA Ames Research Center, Mountain View, CA, United States; and the University of Colorado Boulder, Boulder, CO, United States.

This manuscript was received for review in February 2023. It was accepted for publication in September 2023.

Address correspondence to: James A. Kole, Associate Professor, School of Psychological Sciences, University of Northern Colorado College of Education and Behavioral Sciences, Campus Box 94, Greeley, CO 80639, United States; james.kole@unco.edu.

DOI: https://doi.org/10.3357/AMHP.6250.2023

more than 18,300 applicants to the astronaut training program,[12] with less than 0.1% accepted into the program.

These requirements were established to ensure that selected candidates successfully complete the astronaut training program. Those who are ultimately selected begin a 2-yr training program to acquire the knowledge and skills required for space missions. The requisite learning includes some physical/motor skills, such as spacewalking, as well as complex cognitive tasks, such as foreign language learning and International Space Station operations.[12] Missions are generally between 3 and 6 mo. Thus, training future astronauts involves the following issues: 1) how to retain both simple and complex knowledge and skills over extended time periods (during training, between training and subsequent missions, and during missions); 2) how to ensure that knowledge acquired in one physical context (Earth-based training) is available in other contexts (space-based missions); and 3) how to ensure knowledge and skills may be used creatively in novel situations; that is, how to ensure transfer of learning.

Previous research has established many training principles to maximize the retention and transfer[7] of both procedural (physical/motor skills) and declarative (factual) knowledge.[10] For example, by the Procedural Reinstatement Principle, for simple motor tasks, retention and transfer are maximized when training and test procedures match. By the Variability of Practice Principle, for both procedural and declarative tasks, increasing learning set size (the number of learned exemplars) and randomizing the order of learned exemplars increase both retention and transfer. Such evidence-based training principles might be considered when designing astronaut training programs. However, it is also widely noted that most research in psychology uses undergraduates, often enrolled in introductory psychology classes, as subjects. Arnett[1] found that in one flagship journal, approximately 67% of American psychology studies used undergraduate samples; other journals report similar percentages.[14] These undergraduate samples differ from the general population in terms of personality and attitudinal variables[6] as well as in the magnitude and direction of some experimental manipulations.[13] More relevant to astronaut training, given the strict requirements and low acceptance rate of the training program, undergraduate samples might also differ from astronaut candidates in terms of aptitude, motivation, interests, and physical abilities, and these differences might impact their ability to retain knowledge over time and to transfer learning to new situations. However, it is also possible that presumed differences between astronaut candidates and undergraduate students have lessened due to new, STEM focused subdisciplines within psychology, such as neuroscience. Although not a comparison of astronaut candidates to undergraduates, previous research has shown that there are aptitude differences between those selected and not selected for astronaut training.[3]

The literature on transfer of learning is vast; however, it is only recently that individual differences in the ability to transfer have been examined. For example, McDaniel et al. found that individuals who form more abstract task-based representations transfer learning to a greater extent than individuals who do not form such representations, and the ability to do so is related to working memory (WM) capacity.[11]

Prior to the study by McDaniel et al.,[11] two studies conducted almost a century ago examined individual differences in aptitude (intelligence) and how they relate to transfer, with mixed results. In a study of elementary school children, Brooks[5] found that those with higher aptitude were better able to transfer learning of mental multiplication to mental division. In contrast, Ryans[15] found that higher aptitude led to negative transfer among high school students. Ryans[15] used a digit-letter substitution task whereby letters were associated with digits (e.g., H-1, F-2) and subjects wrote the associated numbers for a series of letters. Subjects trained on one set of letter-number associations and transfer was tested by changing letter-number associations (e.g., H-5, F-4). A negative transfer effect was found whereby performance on the transfer test was lower than performance during training, with the higher aptitude students showing greater negative transfer than others. This finding suggests the higher aptitude group were better at learning the letter-number associations presented during training, which consequently interfered more with the transfer task. Using an undergraduate student sample, Vlach and Kalish showed that initial learning (i.e., performance during study) can moderate the degree of transfer, whether positively or negatively (e.g., when learned responses support or interfere with acquiring new responses).[18] Given that aptitude differences might exist between astronaut candidates and undergraduates, these classic studies, although with school- and college-age samples, suggest that training principles designed to maximize retention and transfer might not apply fully to astronaut training if these principles were derived from studies using undergraduates. One of the purposes of the present study is to examine this possibility.

Also relevant to astronaut training is cognitive load, defined as the demands on WM imposed by a task. In the original formulation of Cognitive Load Theory, a task's demands could be differentiated into three different types of cognitive load: intrinsic, extraneous, and germane.[2] Intrinsic load, namely that which is intrinsic to the task, has to do with the inherent difficulty or complexity of the task being performed or the information being learned, with more difficult/complex tasks and concepts resulting in greater load. Extraneous load is caused by factors external to the task being performed or information being learned that also require WM resources. For example, in a dual-task paradigm whereby two tasks are performed simultaneously, one task may be perceived as creating an extraneous cognitive load for the other. Lastly, germane load refers to the WM requirements necessary to develop task-specific schema in long-term memory that facilitate performance in the future.[16] For example, variability of practice increases germane load because increasing the number of learned exemplars and randomizing the order of exemplars both increase WM demands and make the task more difficult, but also aid in forming task-specific schema. Increasing germane load, as well as reducing extraneous load, during learning has been found to increase transfer during testing in a problem-solving task.[17] However, less explored is whether or not cognitive load during testing influences transfer performance; it has been common to vary cognitive load during learning, but less common to vary cognitive load during testing to see whether such a variation influences the degree of transfer.

The present study was designed to examine three specific issues. The first was to evaluate astronaut candidate performance in terms of retention and transfer on a simple perceptual-motor task as well as on a complex memory task. The second was to compare astronaut candidates and undergraduate subjects to determine if they differ in terms of patterns of retention and transfer, given the high selection criteria for the astronaut training program. Lastly, a third set of subjects (candidate-like) was also included; candidate-like subjects were similar to astronaut candidate subjects in age, education, and general health. Using astronauts regularly in training studies is not feasible due to time and cost constraints. Thus, the purpose of including a candidate-like sample was to determine if they provide a better approximation of astronaut candidate performance in terms of retention and transfer than do undergraduate subjects.

Learning both simple and complex tasks is required of training for missions, and aptitude might impact transfer of previous learning to novel situations, whether positively[5] or negatively.[15] The present study also examined the effect of adding cognitive load to simple and complex tasks, as well as whether the addition of such load differentially affects astronaut candidates, candidate-like, and undergraduate samples. The simple[8] and complex[9] tasks employed in this study were used in several previous studies, to develop training principles using undergraduate samples, and the timing schedule of training and test sessions is based on actual schedules used during space missions.

METHODS

The purpose of Experiment 1 was to compare astronaut candidate, candidate-like, and undergraduate subjects on the simple perceptual-motor task of data entry, in which subjects are presented four-digit numbers on a computer screen and use their dominant hand to type the numbers using the keypad of a keyboard. All subjects were trained in this task, with transfer and cognitive load manipulations introduced at later test sessions. The experiment was conducted over a 500-d period, which allowed for the examination of long-term retention and transfer.

The purpose of Experiment 2 was to compare astronaut candidate, candidate-like, and undergraduate subjects on a complex task. The continuous memory-updating task involves learning name-location associations and both prospective and retrospective memory.[9] For this task, subjects are told that they are tracking the locations of different crewmembers on a planetary surface and, in most cases, have to report the location later by clicking on a map labeled Spacecraft, but in special cases report the location later using a map labeled Mission Control. During study trials subjects had to update their memory for the most recent location of a given name; one of eight names (Alpha, Bravo, etc.) was presented along with one of four locations (East, West, etc.) and subjects were to click the appropriate location on the Spacecraft map. During most study trials, the name-location was presented in black (retrospective) and in a smaller subset of special study trials, the name-location was presented in green (prospective). In test trials, a name was presented in blue and without a location; subjects were to click on the last location associated with that name. In retrospective trials, subjects used the Spacecraft map to recall the last location and on (special) prospective trials, they used instead the Mission Control map. Thus, two different measures of memory were assessed during test trials: memory for the correct location (North, South, East, West) and memory for the correct map (Spacecraft, Mission Control).

Also, two different retention intervals between study and test trials were compared in Experiment 2 to evaluate the extent of forgetting the name-location associations. One interval was shorter, with only a single interpolated trial (2-back), which relies primarily on short-term memory, and the other interval was longer, with seven interpolated trials (8-back), which relies primarily on long-term memory. Thus, this experimental paradigm included manipulations of short- vs. long-term memory, as well as prospective vs. retrospective memory.

Subjects

In Experiment 1 (data entry task), 26 undergraduates, most majoring in STEM fields at the University of Colorado Boulder, participated for compensation (see **Table I** for declared majors of undergraduate subjects). In addition, 20 compensated subjects were recruited from Johnson Space Center's test subject database: 11 astronaut candidates and 9 candidate-like subjects.

Table I. College Majors* and Year in School for Undergraduate Subjects.

COLLEGE MAJOR	N	YEAR IN SCHOOL: FRESHMAN	SOPHOMORE	JUNIOR	SENIOR
Engineering					
Aerospace Engineering	8	7	1	0	0
Computer Science	1	1	0	0	0
Mechanical Engineering	1	1	0	0	0
Pre-Engineering	1	1	0	0	0
Natural Sciences					
Astrophysics	2	1	0	1	0
Biochemistry	1	0	1	0	0
Biology	4	1	2	0	1
Chemistry	1	0	1	0	0
Integrative Physiology	3	1	0	2	0
Neuroscience	5	1	2	2	0
Physics	2	2	0	0	0
Non-STEM					
Communications	1	0	0	1	0
Marketing	1	1	0	0	0
Political Science	1	0	0	1	0

*Six subjects had double majors.

Table II. Subject Demographics for Astronaut Candidate, Candidate-Like, and Undergraduate Subjects.

		HANDEDNESS		SEX		
GROUP	*N*	RIGHT	LEFT	FEMALE	MALE	AGE RANGE (yr)
Astronaut Candidate	11	10	1	5	6	30-55
Candidate-Like	9	7	2	5	4	25-55
Undergraduate	26	25	1	11	15	19-35

The 11 astronaut candidates were in the process of completing astronaut training. (All 11 subsequently completed their training and were eligible to be assigned to missions.) The nine candidate-like subjects met similar educational and physical requirements as astronaut candidates (see **Table II** for demographics of all subject groups). Specifically, candidate-like subjects held degrees in STEM fields and were restricted in age from 25 to 55 yr old, in vision to normal or corrected-to-normal, in height from 58.5″ to 76″, and in self-reported BMI to less than 25.

The same 46 subjects participated in Experiment 2 (continuous memory updating task). Both Experiments 1 and 2 were approved by the Institutional Review Boards at the University of Colorado Boulder and at the National Aeronautics and Space Administration. Informed consent was obtained before subjects participated in the study.

Procedure

In Experiment 1, for the standard data entry task, subjects were presented with a four-digit number on the computer screen (e.g., 4368) and typed the number using their dominant hand. Within each four-digit number, no digit was repeated and 0 was not included. Each number was presented individually, and subjects were instructed to type as quickly and as accurately as possible. For the nondominant variant of the task, subjects were presented four-digit numbers but typed the numbers using their nondominant hand instead of their dominant hand. For the code variant, subjects were presented with four letters and typed their corresponding alphabetical positions (e.g., for dcfh, they typed 4368) using their dominant hand.

In a preliminary session for Experiment 1, subjects were familiarized with the standard data entry task as well as the two variants used in this experiment. Subjects completed 50 trials of the standard, nondominant, and code tasks in that order.

Following this preliminary session were two training sessions. The first training session occurred approximately 88 d (M = 87.65 d, SD = 3.91 d) after the preliminary session, and the second training session occurred approximately 98 d (M = 97.5 d, SD = 8.31 d) after the first training session. For both training sessions, subjects completed three blocks of the standard data entry task. Each block included 100 trials. The same 100 4-digit numbers were presented once in each block of training; these 100 numbers were different from those used during the preliminary session.

Following training were two test sessions in Experiment 1. The first test session occurred approximately 189 d (M = 189.37 d, SD = 11.29 d) after the second training session. The first test session included 100 trials of standard data entry followed by 100 trials of the nondominant variant. Of the 100 trials for each task (standard, nondominant), 50 were old (i.e., presented one time in each training block). The remaining 50 trials were new (i.e., not presented in training). The new trials were also four-digit numbers in which no digit was repeated and the digit 0 was not used. Thus, the standard task examines retention and the nondominant variant examines transfer. More specifically, the nondominant variant examines cognitive transfer because the motor components of the task differ (e.g., using the left hand instead of the right) for the standard and nondominant tasks, but the cognitive requirements (i.e., perceptually processing four digits) are the same.

The second test session of Experiment 1 occurred approximately 64 d (M = 64.04 d, SD = 5.10 d) after the first test session. The second test session included 100 trials of standard data entry followed by 100 trials of the code variant. Of the 100 trials for each task (standard, code), 50 were old trials and the other 50 were new trials, which were different from those used during the first test session. The code variant examines motor transfer because the cognitive components of the task are different for the standard and code tasks (i.e., perceptually processing numbers instead of letters), but the motor requirements are the same. This variant also examines the effect of increasing intrinsic cognitive load on transfer due to the increase in processing demands.

A concluding session for Experiment 1, approximately 62 d (M = 61.93 d, SD = 9.53 d) after the second test session, was the same as the preliminary session. (For a summary of the timing of sessions as well as the tasks completed within each session, see **Table III**.)

All sessions in Experiment 1 took place within a window of approximately 500 d (M = 500.18 d, SD = 9.83 d). Because this study was originally scheduled to be conducted with astronauts while onboard the International Space Station, the training schedule was set to fit their preparation for a standard duration mission of approximately 6 mo, and the testing schedule was set to fit their time in orbit. Ultimately, this study

Table III. Data Entry and Memory Updating Task Variants Completed During Each Session.

		TASK	
SESSION	DAY	DATA ENTRY	MEMORY UPDATING
Preliminary	0	Standard (50) Nondominant (50) Code (50)	Silence (37) Secondary Task (37)
Training 1	88	Standard (300)	Silence (124)
Training 2	185	Standard (300)	Silence (124)
Test 1	375	Standard (100) Nondominant (100)	Silence (124)
Test 2	439	Standard (100) Code (100)	Secondary Task (124)
Concluding	500	Standard (50) Nondominant (50) Code (50)	Silence (37) Secondary Task (37)

Number of trials of each task is indicated in parentheses.

was conducted on the ground only; however, the same schedule was used.

The two tests in Experiment 1 were always in the same order (the nondominant variant was included during the first test session, and the code during the second test session). Also, within each test session, the standard task always preceded the novel variant. Other confounding variables necessarily differentiate the standard task from the variants (e.g., only the standard task was used in training and was simpler than the variants). These confounding variables should be taken into account when interpreting the results, but should not affect the most crucial comparison of undergraduate to astronaut candidate and candidate-like subjects, as the same confounds existed for all three groups.

Experiment 2 (continuous memory updating task) was conducted concurrently with Experiment 1. For each of the six sessions, the memory updating task was completed by all subjects directly after the data entry task. Thus, the number and timing of sessions for Experiment 2 was identical to that for Experiment 1.

In a preliminary session for Experiment 2, subjects were familiarized with and practiced two variants of the memory updating task. For the first variant, subjects performed the memory updating task in silence, whereas for the second, they also performed a concurrent secondary task requiring them to count aloud backward from 100 by 3's. The additional WM resources demanded by the counting backward task increased extraneous load on the updating task. Subjects completed 37 trials in silence and 37 trials with the concurrent secondary task; in each case, the 37 trials included 25 study trials, 4 prospective test trials (2 2-back and 2 8-back), and 8 retrospective test trials (4 2-back and 4 8-back).

The next two sessions in Experiment 2 involved training. Two identical blocks of 124 trials were used in each training session. The 124 trials included 76 study trials, 16 prospective test trials (8 2-back and 8 8-back), and 32 retrospective test trials (16 2-back and 16 8-back).

Transfer testing for Experiment 2 occurred during the fourth and fifth sessions. There were two transfer tests (standard and counting), both involving different stimuli from those shown during training. The first transfer test, during the fourth session, required no secondary task. The second transfer test, during the fifth session, required the concurrent secondary counting backward task. Only the order of the names (and hence the name-location associations) differed between the four training blocks and the two test blocks, each of which included 124 trials.

The concluding session of Experiment 2 was the same as the preliminary session. (See Table III for a summary of tasks completed within each session.)

Statistical Analysis

For both Experiments 1 and 2, all statistical tests were mixed-factorial analyses of variance with $\alpha = 0.05$.

For Experiment 1, analyses of training and each test included the between-subjects variable of subject group (astronaut candidate, candidate-like, undergraduate) coupled with two within-subject independent variables. The within-subject variables for training were training session (1 or 2) and block (1, 2, or 3). The within-subject variables for the first test were task (standard, nondominant) and trial type (new, old), and for the second test were task (standard, code) and trial type. In each case, the dependent variables examined were accuracy as well as total response time (TRT; the time required to type the four digits and the concluding Enter key from stimulus onset) for correct trials (i.e., trials in which all four digits were typed correctly).

For Experiment 2, training data were averaged over the four blocks and training analyses included one between-subjects variable, subject group (astronaut candidate, candidate-like, undergraduate), and two within-subject variables, trial type (retrospective, prospective) and retention interval (2-back, 8-back). Test analyses also included the between-subjects variable of subject group and the within-subject variables of trial type and retention interval, with the additional within-subject variable of test (Test 1, Test 2). There were two dependent variables: proportion correct location (North, South, East, West), which assessed memory for the location associated with a given name, and proportion correct map (Spacecraft, Mission Control), which assessed whether subjects used the correct map to report location.

RESULTS

For brevity and clarity, we report all significant main effects and interactions involving the variable of subject group, as well as other effects and interactions of theoretical interest. Analysis of Experiment 1 training accuracy revealed differences between subject groups; astronaut candidate ($M = 0.959$) and candidate-like subjects ($M = 0.959$) demonstrated higher accuracy than undergraduates ($M = 0.926$) [$F(2, 41) = 3.566$, MSE = 0.010, $P = 0.037$, $\eta_p^2 = 0.148$].

For Experiment 1 training TRT, average TRT decreased from the first training session (M = 2.891 s) to the second (M = 2.831 s) [$F(1, 41) = 4.895$, MSE = 0.033, $P = 0.033$, $\eta_p^2 = 0.107$], as well as across blocks within each training session (Block 1 = 2.881 s; Block 2 = 2.864 s; Block 3 = 2.838 s) [$F(2, 82) = 4.179$, MSE = 0.010, $P = 0.019$, $\eta_p^2 = 0.092$]. There was a marginally significant difference in TRT between subject groups, with faster TRT for astronaut candidates (M = 2.664 s) than for candidate-like (M = 2.798 s) and undergraduate subjects (M = 2.963 s) [$F(2, 41) = 3.048$, MSE = 0.671, $P = 0.058$, $\eta_p^2 = 0.129$]. A post hoc test revealed that the difference between astronaut candidate and candidate-like subjects (combined) and undergraduates was significant [$F(1, 42) = 5.368$, MSE = 0.667, $P = 0.026$, $\eta_p^2 = 0.113$].

Analysis of Experiment 1 Test 1 accuracy showed performance on the standard task was higher for old than for new numbers for astronaut candidate and candidate-like subjects, which reflects retention, whereas the opposite pattern was observed for undergraduates. For the nondominant variant, which involved transfer, the reverse pattern was observed: higher accuracy for new than for old numbers for astronaut candidate and candidate-like subjects, but higher accuracy for old than for new numbers for undergraduates [$F(2, 43) = 4.466$, MSE = 0.001, $P = 0.017$, $\eta_p^2 = 0.172$ (see **Fig. 1**)].

For Experiment 1 Test 1 TRT, averaging across tasks, old numbers (M = 2.787 s) were entered more quickly than new numbers (M = 2.839 s) [$F(1, 43) = 25.532$, MSE = 0.004, $P < 0.001$, $\eta_p^2 = 0.373$], documenting retention of numbers practiced at training.

For Experiment 1 Test 2, accuracy was higher for astronaut candidate (M = 0.910) and candidate-like (M = 0.924) subjects than for undergraduates (M = 0.857) [$F(2, 43) = 4.392$, MSE = 0.019, $P = 0.018$, $\eta_p^2 = 0.170$]. However, there was a significant interaction between subject group and task [$F(2, 43) = 4.337$, MSE = 0.007, $P = 0.019$, $\eta_p^2 = 0.168$]. For the standard task, which involved lower intrinsic load, accuracy was similar for the three subject groups; in contrast, for the code task, which involved higher intrinsic load, accuracy was higher for astronaut candidate and candidate-like subjects than for undergraduates (see **Fig. 2**).

For the analysis of Experiment 1 Test 2 TRT, old numbers (M = 4.072 s) were entered more quickly than new numbers averaging across tasks (M = 4.439 s) [$F(1, 43) = 133.616$, MSE = 0.035, $P < 0.001$, $\eta_p^2 = 0.757$].

For Experiment 2 location accuracy at training, performance on prospective trials (M = 0.664) was better than on retrospective trials (M = 0.517) [$F(1, 41) = 35.875$, MSE = 0.082, $P < 0.001$, $\eta_p^2 = 0.467$]. Thus, when subjects had to use a different map than usual to recall locations, they were more accurate on locations than when they used the standard map. Performance was also better at the short (2-back) retention interval (M = 0.698) than at the long (8-back) retention interval (M = 0.483) [$F(1, 41) = 122.065$, $MSE = 0.059$, $P < 0.001$, $\eta_p^2 = 0.749$]. The effect of retention interval was much larger for retrospective than for prospective trials [$F(1, 41) = 22.211$, MSE = 0.024, $P < 0.001$, $\eta_p^2 = 0.351$], supporting the protective function of special responses.[4] That is, memory for locations decreased as the number of intervening trials increased; however, this decrease was greater on retrospective trials (2-back, M = 0.654; 8-back, M = 0.379) than on prospective trials (2-back, M = 0.741; 8-back, M = 0.587). There were also subject group differences, with accuracy higher for astronaut candidate ($M = 0.691$) and candidate-like ($M = 0.632$) subjects than for undergraduates ($M = 0.535$) [$F(2, 41) = 8.402$, MSE = 0.184, $P < 0.001$, $\eta_p^2 = 0.291$].

For the analysis of Experiment 2 map accuracy at training, there were subject group differences with accuracy higher for astronaut candidate (M = 0.768) and candidate-like (M = 0.718) subjects than for undergraduates (M = 0.634) [$F(2, 41) = 4.919$, MSE = 0.192, $P = 0.012$, $\eta_p^2 = 0.194$]. For Experiment 2, location accuracy was higher for Test 1 (M = 0.608) than for Test 2 (M = 0.386), during which there was a concurrent counting backward task [$F(1, 43) = 59.747$, MSE = 0.052, $P < 0.001$, $\eta_p^2 = 0.581$]. Location accuracy was also higher for astronaut candidate (M = 0.599) and candidate-like (M = 0.551) subjects than for undergraduates (M = 0.435) [$F(2, 43) = 10.059$, MSE = 0.096, $P < 0.001$, $\eta_p^2 = 0.319$]. However, the differences between subject groups depended on test session and trial type [$F(2, 43) = 4.346$, MSE = 0.022, $P = 0.019$, $\eta_p^2 = 0.168$]. For Test 1, location accuracy was higher on prospective than on retrospective memory trials for all subject groups. For Test 2 the same pattern was evident, but weaker for undergraduates, most likely because performance was approaching floor (chance level, 0.25) on both trial types (see **Fig. 3**).

Experiment 2 map accuracy at test also showed that accuracy was higher for astronaut candidate (M = 0.725) and candidate-like (M = 0.707) subjects than for undergraduates (M = 0.587) [$F(2, 43) = 10.959$, MSE = 0.071, $P < 0.001$, $\eta_p^2 = 0.338$]. Map accuracy was also higher for the first (M = 0.708) than for the second (M = 0.579) test, which involved the concurrent secondary task [$F(1, 43) = 43.096$, MSE = 0.025, $P < 0.001$, $\eta_p^2 = 0.501$]. The three-way interaction between subject group, test, and trial type was also significant [$F(2, 43) = 5.702$, MSE = 0.045, $P = 0.006$, $\eta_p^2 = 0.210$] because during Test 1, map accuracy was higher on retrospective than on

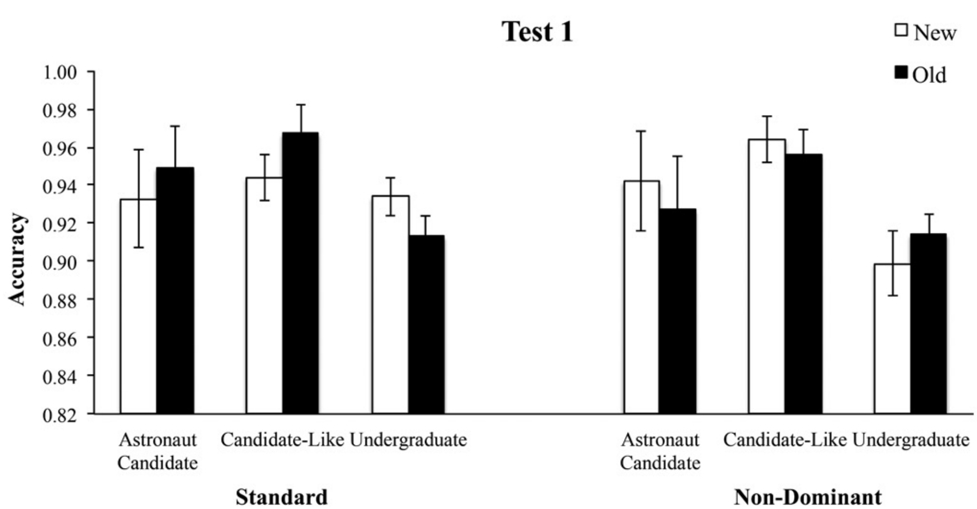

Fig. 1. Accuracy as a function of subject group, task, and trial type during Test 1 in Experiment 1. Error bars represent between-subjects standard errors of the mean.

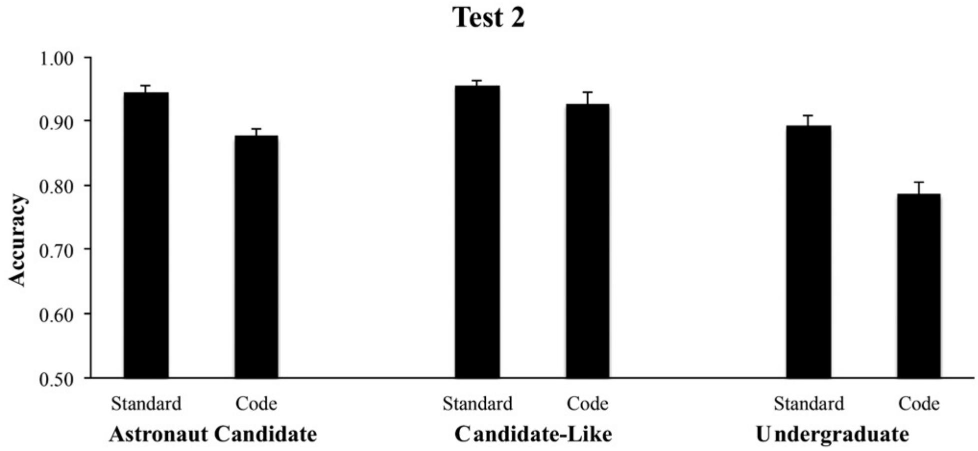

Fig. 2. Accuracy as a function of subject group and task during Test 2 in Experiment 1. Error bars represent between-subjects standard errors of the mean.

prospective trials for all subject groups, likely due to response bias because there were more trials requiring the Spacecraft map (study and retrospective test trials) than the Mission Control map (only prospective test trials). For Test 2, the same pattern was evident, but was especially strong for undergraduates (see **Fig. 4**). Undergraduates were particularly impacted negatively by the concurrent secondary task on prospective memory trials.

DISCUSSION

The present study compared astronaut candidates and candidate-like subjects to undergraduates on a simple motor task as well as on a more demanding complex memory task. More specifically, for each task, subject groups were compared on retention and transfer as well as under conditions of lower and higher cognitive load.

For both tasks, there were overall differences in performance between astronaut candidate, candidate-like, and undergraduate subjects, with the former two subject groups performing better on measures of speed and accuracy. For the simpler data entry task, η_p^2 for main effects of subject group and interactions involving subject group indicated a medium to large effect size, and for the more complex memory updating task, η_p^2 indicated a large effect size. Thus, differences between subject groups were more pronounced with increased task complexity. Astronaut candidate and candidate-like subjects were also less impacted by intrinsic cognitive load (Experiment 1) and extraneous cognitive load (Experiment 2). Given the cognitive loads experienced and expected during space missions, especially when dealing with anomalies, this finding is particularly important.

For the data entry task in Experiment 1, different patterns of retention and transfer were found across subject groups: astronaut candidate and candidate-like subjects showed greater retention of training stimuli than undergraduate subjects; however, they also showed a negative transfer effect such that when typing training stimuli using their nondominant hand, they were less accurate than when typing entirely new numbers. The opposite pattern was found for undergraduates. The finding of a negative transfer effect for astronaut candidate and candidate-like subjects is similar to that found by Ryans[15] for higher aptitude subjects (who were high school students). It is

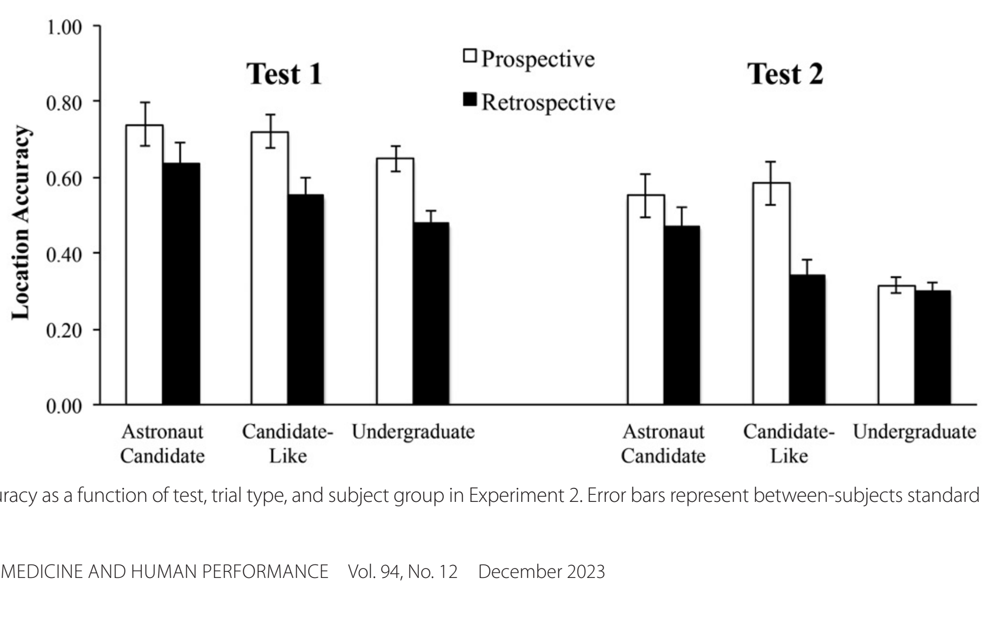

Fig. 3. Location accuracy as a function of test, trial type, and subject group in Experiment 2. Error bars represent between-subjects standard errors of the mean.

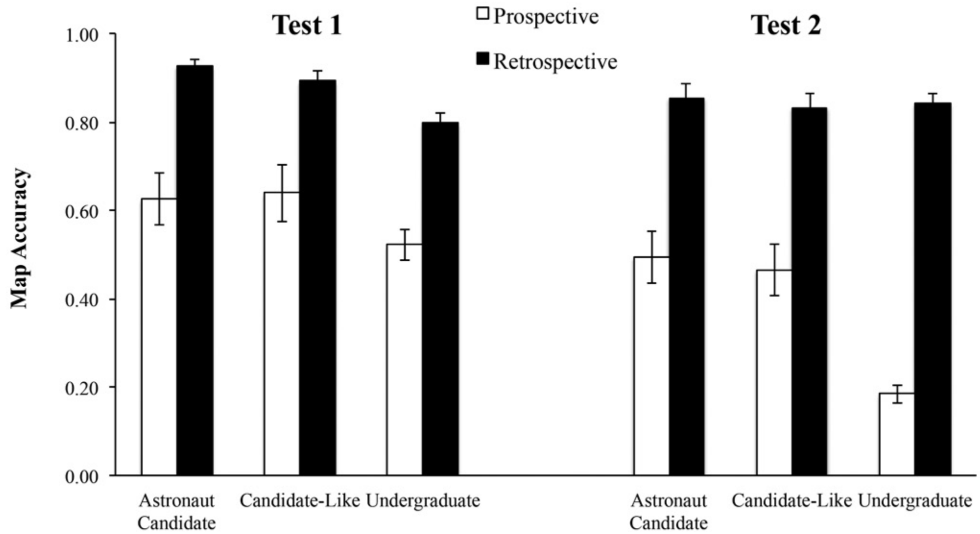

Fig. 4. Map accuracy as a function of test, trial type, and subject group in Experiment 2. Error bars represent between-subjects standard errors of the mean.

possible that astronaut candidate and candidate-like subjects learned the associated response patterns for each four-digit number better than undergraduates, which in turn facilitated performance for the standard task but hindered performance for the nondominant variant when a different response pattern was required. These results also suggest a tradeoff between retention and transfer and are similar to other studies showing that the forgetting of specific training stimuli can enhance generalization and transfer.[18] The finding that high retention might interfere with transfer is also of particular interest. Given the limited ability to predict events during future missions, the transfer of trained skill and knowledge might be more important than just sheer retention. Astronauts will have to manage the unexpected, and they will have to do that with the skill and knowledge they acquired during training. Thus, such training might need to be focused on optimizing transfer rather than emphasizing retention.

The present study suggests that research with candidate-like subjects can better predict astronaut candidate performance than research with undergraduates, at least on Earth. There is still, however, a big gap in our understanding of performance during space missions. There is anecdotal evidence from astronauts' stories about being "space stupid," and from anecdotal reports that many trained tasks require significantly more time in space than on Earth. But so far, no systematic collection of astronauts' retention and transfer performance data has been completed in space.

As mentioned earlier, the study reported here was originally designed to be conducted onboard the International Space Station to provide the very data needed to understand the extent to which results obtained in Earth-based studies can indeed predict performance in space, but unfortunately the astronauts ultimately did not perform the studies in space. Further, it is not known whether astronaut performance during a standard duration mission can accurately predict performance during a long-duration, exploration-class mission such as a crewed mission to Mars. To best design training for future missions, such that retention and transfer are optimized, especially transfer for unexpected situations of high cognitive load and complex tasks of high intrinsic and extraneous load, studies will have to be done that approximate as much as possible the conditions expected for such future missions.

ACKNOWLEDGMENTS

Many thanks to Donna Dempsey for help recruiting astronaut candidate and candidate-like subjects from Johnson Space Center and for collecting data from these subjects. Results involving university undergraduates were previously reported in two journal articles. Preparation of this paper occurred while Alice Healy was a visiting scholar in the Computational Memory Lab at the University of Pennsylvania.

Financial Disclosure Statement: This research was supported in part by National Aeronautics and Space Administration Grants NCC2-1310 to the University of Colorado Boulder and NNX17AE07A to the University of Northern Colorado. The authors have no competing interests to declare.

Authors and Affiliations: James A. Kole, M.A., Ph.D., University of Northern Colorado, Greeley, CO, United States; Immanuel Barshi, Ph.D., Ph.D., NASA Ames Research Center, Mountain View, CA, United States; and Alice F. Healy, A.B., Ph.D., and Vivian I. Schneider, M.A., Ph.D., University of Colorado Boulder, Boulder, CO, United States.

REFERENCES

1. Arnett JJ. The neglected 95%: why American psychology needs to become less American. Am Psychol. 2008; 63(7):602–614.
2. Bannert M. Managing cognitive load—recent trends in cognitive load theory. Learn Instr. 2002; 12(1):139–146.
3. Bishop SL, Faulk D, Santy PA. The use of IQ assessment in astronaut screening and evaluation. Aviat Space Environ Med. 1996; 67(12):1130–1137.

4. Bourne LE Jr, Healy AF, Bonk WJ, Buck-Gengler CJ. Intention to respond in a special way offers some protection against forgetting associations. Am J Psychol. 2011; 124(1):23–36.
5. Brooks FD. The transfer of training in relation to intelligence. J Educ Psychol. 1924; 15(7):413–422.
6. Hanel PH, Vione KC. Do student samples provide an accurate estimate of the general public? PLoS One. 2016; 11(12):e0168354.
7. Healy AF, Kole JA, Bourne LE. Training principles to advance expertise. Front Psychol. 2014; 5:131.
8. Healy AF, Kole JA, Schneider VI, Barshi I. Training, retention, and transfer of data entry perceptual and motor processes over short and long retention intervals. Mem Cognit. 2019; 47(8):1606–1618.
9. Healy AF, Schneider VI, Buck-Gengler CJ, Kole JA, Barshi I. Intention to respond in a special way protects against forgetting associations even when working memory is occupied. Am J Psychol. 2020; 133(4):415–426.
10. Kole JA, Healy AF, Schneider VI, Barshi I. Training principles for declarative and procedural tasks. In: Landon LB, Slack KJ, Salas E, editors. Psychology and human performance in space programs: research at the frontier. Abingdon (UK): Taylor & Francis; 2021:131–149.
11. McDaniel MA, Cahill MJ, Robbins M, Wiener C. Individual differences in learning and transfer: stable tendencies for learning exemplars versus abstracting rules. J Exp Psychol Gen. 2014; 143(2):668–693.
12. National Aeronautics and Space Administration. Astronaut requirements. NASA; 2020. [Accessed 2023 Oct. 12]. Available from: https://www.nasa.gov/audience/forstudents/postsecondary/features/F_Astronaut_Requirements.html.
13. Peterson RA. On the use of college students in social science research: insights from a second-order meta-analysis. J Consum Res. 2001; 28(3): 450–461.
14. Pollet TV, Saxton TK. How diverse are samples used in the journals "Evolution & Human Behavior" and "Evolutionary Psychology"? Evol Psychol Sci. 2019; 5(3):357–368.
15. Ryans DG. An experimental study of the transfer of training with special attention to the relation of intelligence test performance. J Educ Psychol. 1936; 27(7):492–500.
16. van Gog T, Paas F, Sweller J. Cognitive load theory: advances in research on worked examples, animations, and cognitive load measurement. Educ Psychol Rev. 2010; 22(4):375–378.
17. van Merriënboer JJG, Schuurman JG, de Croock MBM, Paas FGWC. Redirecting learners' attention during training: effects on cognitive load, transfer test performance and training efficiency. Learn Instr. 2002; 12(1): 11–37.
18. Vlach HA, Kalish CW. Temporal dynamics of categorization: forgetting as the basis of abstraction and generalization. Front Psychol. 2014; 5:1021.