Journal of Experimental Psychology: Applied
1996, Vol. 2, No. 3, 227-249

Analysis of Part-Task Training Using the Backward-Transfer Technique

Barry P. Goettl and Valerie J. Shute
Armstrong Laboratory, U.S. Air Force

Researchers conducted 2 experiments that used backward transfer to improve the efficiency of part-task training for a desktop flight simulator. In Experiment 1, a part-task group showed positive transfer but did not perform as well as a whole-task group. Backward-transfer analysis indicated that only a subset of the component tasks was critical to the criterion task. In Experiment 2, a part-task training regime that used the critical component tasks was compared with a whole-task regime and a part-task regime composed of noncritical component tasks. Results indicated that the critical part-task regime was as effective as the whole-task regime, validating the utility of the backward-transfer technique.

The research discussed in this article examined complex skill acquisition involving a dynamic spatial task, coupled with an analysis of the effects of a backward-transfer approach to training. A discussion of part-task training and the factors that influence the effectiveness of part-task training is followed by an overview of the transfer of training paradigm used to test part-task training effectiveness. We conclude our literature review with an in-depth discussion of the specific training technique used in our two experimental studies; namely, the backward-transfer approach.

Part-Task Training

Part-task training has long been a popular approach to training complex manual control and tracking tasks. The basic tenet underlying part-task training is that drill on individual components of a complex skill will improve performance on the complex skill. Carlson, Khoo, and Elliot (1990) refer to this hypothesis as the *component fluency hypothesis* (p. 267). Carlson, Sullivan, and Schneider (1989) point out that this hypothesis rests on three assumptions generally agreed on by theories of cognitive skill: (a) complex skills consist of a hierarchy of basic component skills and organizing strategies; (b) there are capacity limits placed on cognitive processing; and (c) fluency on the component skills is critical to skilled performance on the complex task. Thus, part-task training is thought to permit individuals to acquire critical component skills that transfer to the whole task without imposing the cognitive demands of the whole task.

In spite of the intuitive appeal and theoretical foundations of the component fluency hypothesis, part-task training has garnered, at best, only modest empirical support and has frequently been less effective than whole-task training (for a review, see Wightman & Lintern, 1985). We discuss four challenges that a part-task training regime must overcome to be effective. First, part-task training effectiveness depends on the identification of valid critical component tasks. Second, the skills identified as most critical early in training may not be critical later in training. Third, interactions among the component tasks play an important role in the whole task. Finally, individual differences in ability and style of learning play a large role in skill

Barry P. Goettl and Valerie J. Shute, Armstrong Laboratory, Human Resource Directorate, Technical Training Research Division, Intelligent Training Branch (AL/HRTI), Brooks Air Force Base (AFB), U.S. Air Force.

Funds for this research were provided by the U.S. Air Force Office of Scientific Research and the Training Research for Automated Instruction (TRAIN) Project of AL/HRTI at Lackland AFB under the direction of J. Wesley Regian. The opinions expressed in this article are those of the authors and do not necessarily reflect those of the U.S. Air Force.

Correspondence concerning this article should be addressed to Barry P. Goettl, AL/HRTI, 1800 Carswell Avenue, Lackland AFB, Texas 78236-5507. Electronic mail may be sent via Internet to bgoettl@colab.brooks. af.mil.

acquisition. We examine each of these challenges below.

Identification of Critical Skills

One of the problems faced by part-task training is the difficulty in identifying component skills that are critical for proficiency on the whole-task. Wightman and Lintern (1985) identified three general approaches to part-task training and found that the effectiveness of training depends on the approach used. One approach they examined is called *simplification,* in which a difficult task is made easier by adjusting certain characteristics of the task. For example, Briggs and Naylor (1962) trained male undergraduates on tracking tasks having relatively simple control dynamics and transferred them to tracking tasks having more complex control dynamics. They found that training on the simplified tasks produced positive transfer, but transfer was less than 100%. Although some researchers have shown that simplification can produce transfer (Briggs & Naylor, 1962; Briggs & Waters, 1958), Lintern and Gopher (1980) concluded in their review of research on simplification that the technique is not as effective as whole-task training.

Another technique for part-task training is called *segmentation.* In segmentation, tasks are partitioned on the basis of temporal or spatial dimensions. For example, Bailey, Hughes, and Jones (1980) segmented an air-to-ground attack task into components of downwind leg, base leg, roll-in, and final approach and used a backward-chaining procedure. In backward chaining, the final segment in the sequence (e.g., final approach) is trained to criterion first; then the preceding segment is added (e.g., roll-in and final approach) and so on until the full task is "chained" together. Bailey et al. found that the backward chaining approach was more effective than whole-task training. Wightman and Sistrunk (1987) used segmentation to train a carrier landing task. Participants performed segments of the task starting at 2,000 feet (609.6 m) from touchdown, then 4,000 feet (1,219.2 m) from touchdown, and then 6,000 feet (1,828.8 m) from touchdown. Those receiving segmentation training performed better on the criterion task than those receiving equivalent training on the criterion task. Ash and Holding (1990) used segmentation to divide a keyboard task into three segments of eight sequential notes. They also found that segmentation was superior to whole-task training.

The third approach reviewed by Wightman and Lintern (1985) is called *fractionation,* wherein the task is divided into components that are performed simultaneously. For example, Mané, Adams, and Donchin (1989) used a task analysis of Space Fortress (Mané, Coles, Wickens, & Donchin, 1983) to develop a part-task training regime based on fractionation. Their part-task regime was more effective than whole-task training. This approach was also successfully used by Frederiksen and White (1989) and Fabiani et al. (1989). However, other studies have failed to show that part-task training based on fractionation is more effective than whole-task training (Adams, 1960; Stammers, 1980). Wightman and Lintern concluded that fractionation is more effective than whole-task training, only when a systematic procedure is used to decompose the task.

Critical Skills and Skill Acquisition

The challenge of identifying the critical components of a complex task is exacerbated by the fact that the critical skills may change as expertise develops. Ackerman's (1992) theory of skill acquisition proposes that during the early stages of learning, knowledge is declarative and is based largely on working memory and general intelligence. However, later in learning, consistent components become automated, and the importance of working memory and general intelligence becomes attenuated and the importance of perceptual speed increases. Kanfer and Ackerman (1989, 1996) have presented skill acquisition data on an air traffic control task that were consistent with Ackerman's predictions. The point here is that identification of the critical components based on empirical data may depend on the skill level of the individuals being studied. Moreover, changes in the criticality of skills during acquisition suggest that the order in which component tasks are presented may be important.

Interactions Among Component Tasks

Part-task training effectiveness also depends on the interaction among task components and the strategies for organizing the components into the whole. Gopher, Weil, and Siegel (1989) argued that a complex task is best thought of as "an organized set of response schemas" (p. 148) executed and coordinated by high-level schemas or strategies. They developed a training regime that required individuals to shift emphases between the various components (i.e., ship control, mine management, and bonus management) of a complex task (i.e., Space Fortress). The advantage of their approach is that individuals can avoid the cognitive limits of the whole task by focusing on the components as they occur in the context of the whole task.

A production system approach also emphasizes the importance of interactions among the components and their relationship to the whole task (Anderson, 1983). This approach holds that transfer depends on whether identical procedures can be used in the acquisition and transfer contexts. Singley and Anderson (1989) argued that "knowledge comes in declarative form, is used by weak methods to generate problem solutions, and as a byproduct, new productions are formed. The key step is the knowledge compilation process, which produces the domain-specific skill" (p. 50). Declarative knowledge serves as input for the knowledge compilation phase, thus the quality and structure of this information influences the compilation process. According to this view, exposure to the criterion task (i.e., the whole task) would be most influential early in training when it can provide an appropriate template for the development of domain-specific productions during the knowledge compilation phase. Without early exposure to the criterion task, productions are developed outside the context of the criterion, and therefore may not map onto the relevant domain.

In contrast, Schneider and Detweiler (1988) suggested that exposure to the whole task will be most helpful during the middle phase of training after the component task procedures have been compiled to the point that they are stable. They argued that integration of component skills is accomplished through compensatory activities and that these compensatory activities must be based on stable component task procedures in order to be effective. The data of Carlson, Khoo, and Elliot (1990) supported this hypothesis. They concluded that trainees should not be introduced to the target domain until after they have received sufficient practice on the component tasks. The question of when to introduce the target task is beyond the scope of the present article; however, we provided all participants with a brief, early exposure to the target task (i.e., whole task) in the form of a pretest.

Individual Differences

Differences in spatial ability may play a role in the effectiveness of part-task training of complex spatial tasks such as that used in the present study. Spatial skill represents a major individual differences factor distinct from other abilities like verbal, quantitative, or reasoning skill (see Lohman, 1979, for an excellent review of this factor). Effective use of spatial information is one aspect of human cognition and shows up in situations as diverse as navigating through a novel environment to determining the trajectories of approaching objects. One interesting aspect of spatial ability is that there have been countless studies reported in the literature that have revealed significant gender differences (e.g., Linn & Petersen, 1985; Maccoby & Jacklin, 1974; McGee, 1979; Voyer, Voyer, & Bryden, 1995). Invariably, these studies have shown that men perform better on a wide array of small-scale spatial tasks compared with women.

For instance, Linn and Petersen (1985) conducted a meta-analysis on studies reported in an earlier review on this topic (i.e., Maccoby & Jacklin, 1974). They found gender differences in two categories of spatial tests: spatial perception (effect size $M = 0.44, p < .05$) and mental rotation (effect size $M = 0.73$, $p < .05$). Their third category of spatial tests (i.e., spatial visualization) was not significantly different between women and men (effect size $M = 0.13$, $p > .05$). More recently, Voyer, Voyer, and Bryden (1995) performed a very large meta-analysis of studies cited in the Maccoby and Jacklin (1974) review, studies comprising the Linn and Petersen (1985) meta-analysis, as well as studies culled from an extensive literature search through 1993 on the PsycLit database of the American Psychological Association. Specific tasks were included in this meta-analysis only if at least five studies had already been conducted with the task. This was done to allow meaningful tests of homogeneity and reasonable estimates of effect size for each test. Cohen's d was used in the final analysis of 286 studies. The authors found a mean weighted $d = 0.37$ ($z = 2.61$,

$p < .01$), which indicated highly significant gender differences in spatial abilities that favor men.

Spatial aptitude may also interact with training condition to influence skill acquisition. For instance, inefficiencies in a particular training regime may be most evident for low-aptitude learners. There is a large body of research documenting Aptitude × Treatment interactions (ATI) that shows that the effectiveness of training interventions (treatments) depends on the aptitudes of the learners. For example, Wightman and Sistrunk (1987) showed that high-aptitude learners, but not low-aptitude learners, were able to overcome the initial disadvantage of whole-task training in an aircraft simulator landing task. Similarly, Shute (1993) showed that in the context of a flight-engineering tutor, learners with low incoming knowledge but high working-memory capacity benefited from an extended practice environment. Learners with high incoming knowledge and low working-memory capacities benefited from the abbreviated practice environment. Other learners (i.e., high working memory–high knowledge and low working memory–low knowledge) did not show differential learning in either environment. We will pay particular attention to the potential of ATIs in the current study because of the strong differences in spatial aptitudes that have been observed in the literature.

Transfer and Backward Transfer

Of the challenges faced by part-task training described above, we focus on the problem of identifying component tasks that are critical to the whole task. Wightman and Lintern (1985) suggested that to implement part-task training effectively, critical skills must be identified with a componential analysis of the target task. Subsequently, tasks designed to enhance the critical skills must be developed and validated. One validation technique recommended by Wightman and Lintern (1985) is referred to as *backward transfer.* To illustrate the backward-transfer paradigm, we compare it with a more familiar paradigm called *transfer of training* (TOT).

When the TOT paradigm is used to test part-task training effectiveness, the experimental design should include at least one control group that receives training on the whole or criterion task throughout training and at least one experimental group that receives an equivalent amount of training on one or more part tasks. In addition, both groups should also be given an appropriate test on the criterion task. The test phase is often called a *transfer phase,* even though the control group does not transfer to a different task. This experimental design permits calculation of transfer as well as differential transfer. In estimating transfer, the transfer performance of the experimental group (i.e., part-task group) is compared with the performance of the control group (i.e., the whole-task group) during initial training. Differential transfer involves the comparison of the control and experimental groups relative to equal amounts of experience. In other words, the transfer performance of the experimental group is compared with the transfer performance of the control group.

There are many different formulas for measuring transfer and differential transfer, but in general, transfer can be negative, positive, or even exceed 100%. Negative differential transfer indicates that the training regime disrupted or impeded skill development. Positive differential transfer that is less than 100% indicates that some skills were acquired, but part-task training was less efficient than whole-task training. Differential transfer that exceeds 100% indicates that part-task training is more efficient than whole-task training. Although a training regime that produces positive, differential transfer exceeding 100% is ideal, one that produces positive differential transfer below 100% is of value in applied settings where part-task training is sufficiently safer or less expensive than whole-task training (e.g., aircraft simulator training).

When the backward-transfer paradigm is applied to part-task training, the general experimental design is similar to a TOT design, but the goal is to measure transfer from the component tasks to the criterion task. At least one group is trained on the criterion task (i.e., whole-task group), and at least one group is trained on the component tasks (i.e., part-task group) during the training phase. During the transfer phase, both groups are tested on the component tasks rather than the criterion task. Backward transfer to the component tasks can be estimated by comparing the transfer performance of the whole-task group with the initial training performance of the part-task group. Component tasks that show positive backward transfer involve skills presumably acquired by the whole-task group during training and thus are critical to the whole task.

Salthouse and Prill (1983) used a modified backward-transfer technique to identify the component skills critical to a trajectory intersection task. They trained individuals on the trajectory–intersection task and divided them into high- and low-ability groups that were based on performance level. Then they compared the two groups on several component tasks derived from a preliminary model of the trajectory–intersection task. The model assumed that component tasks were executed sequentially. Although their study is a good illustration of the backward-transfer approach, Salthouse and Prill found that component effectiveness was not strongly related to skilled performance. In addition, improvements in trajectory–intersection performance were not accompanied by improvements on the component tasks. They concluded that skilled performance was associated with more effective strategies for executing components rather than skill on the individual components. That is, high- and low-skill performers showed similar skill levels on the component tasks, but high-skill performers showed evidence of executing the component sequence repetitively, and low-ability individuals executed the entire sequence only once.

There were two major goals of the present study: (a) to determine the utility of the backward-transfer technique for developing effective part-task training regimes and (b) to examine the role of spatial ability differences in the acquisition of a complex spatial task. In Experiment 1, we combined the backward-transfer methodology with a TOT design to identify several component tasks critical to a complex flight task. In Experiment 2, a part-task training regime composed of component tasks deemed critical in Experiment 1 was compared with (a) a whole-task training regime and (b) a part-task training regime composed of noncritical component tasks.

The component tasks were developed on the basis of observations of individuals in several pilot studies and were intended to represent a hierarchy of skills. Some of the tasks focused on basic flight skills related to pitching and rolling[1] the plane. In addition, several tasks required integration of the basic skills into more complex maneuvers. These tasks required individuals to change heading and/or altitude. At the next level in the hierarchy were several tasks requiring participants to fly through a single gate. The tasks at the highest level of the hierarchy required participants to locate a gate on a navigational display and then find it and fly through it. These tasks require three-dimensional spatial orientation to translate the "god's eye" view into the perspective "out-of-the-cockpit" view. These tasks were intended to integrate skills used in all the other tasks. In this hierarchy, the gate-aiming tasks represent the part-task method of segmentation, and all other tasks represent fractionation.

Experiment 1

There were three primary goals underlying Experiment 1. First, we were interested in determining the overall effectiveness of a part-task training approach compared with a whole-task approach using a desktop flight simulator as the criterion task. The second goal was to isolate the component tasks and determine their relevance to the criterion task. Finally, we wanted to explore the relationship between spatial aptitude and training approach. The experimental design is a combination of the TOT and backward-transfer paradigms. The whole-task group received in order a training phase on the criterion task, a transfer phase on the criterion task, and then a backward-transfer phase on the component tasks. The part-task training group received a training phase on the component tasks, a transfer phase on the criterion task, and a backward-transfer phase on the component tasks. This design allows us to estimate: (a) transfer from the component tasks to the criterion task by comparing the transfer performance of the part-task group with the initial training performance of the whole-task group; (b) differential transfer by comparing the transfer performance of the part-task group with the transfer performance of the whole-task group; and (c) backward transfer from the whole task to the component tasks by comparing the backward-transfer performance of the whole-task group to the initial training performance of the part-task group. Differential back-

[1] To *pitch and roll* the plane means to rotate the plane around axes of the plane so that the nose moves up or down relative to the tail (i.e., pitching), or the right wing moves up or down relative to the left wing (i.e., rolling).

ward transfer cannot be estimated because the part-task group experiences the component tasks as well as the criterion task prior to the backward-transfer phase.

Assessment of backward transfer is crucial for identifying the relevant and irrelevant component tasks. Elimination of component tasks irrelevant to the criterion task would presumably render the part-task training regime more efficient. Finally, we wanted to examine performance data on this task in relation to differences in spatial ability. Specifically, we hypothesized a main effect of spatial ability and an interaction between training condition and spatial ability where low-aptitude learners would show larger benefits of part-task training than high-aptitude learners. This hypothesis was motivated by findings of Wightman and Sistrunk (1987) that high-aptitude learners, but not low-aptitude learners, were able to overcome the negative effects of whole-task training. Our hypothesis is also based in part on consideration of potential ceiling effects.

Method

Participants

We recruited 42 men and 38 women through local temporary employment agencies in San Antonio, Texas, to take part in the study. Participants were paid about $5.00 an hour. Participants ranged in age from 18 to 30 and reported spending less than 20 hr per week playing computer and video games (*Mdn* = 2.0 h). All participants had a high school diploma or general equivalency diploma (GED), but none had completed a 2- or 4-year college degree. Also, none of the participants had ever flown the flight simulator used in the present study. The participants were assigned randomly to one of two groups (i.e., the whole-task and part-task training). There were 21 men and 19 women in each group; however 2 women in the whole-task group did not complete the study, and their data were excluded from all analyses.

Equipment

The study was conducted at the Armstrong Training Research for Automated Instruction (TRAIN) Laboratory (Lackland Air Force Base, Texas), which contains 30 laboratory stations, each equipped with a Deskpro 486/33L (Compaq), Multisync 4DS monitor (NEC), and a flight stick (CH Products). A desktop flight simulator called Phoenix (Galaxy Scientific, San Antonio, TX) was used for training and data collection. This desktop simulator utilizes simplified dynamics in that the three axes of aircraft rotation are independent. For example, changes in roll do not effect pitch or altitude. The Phoenix display (see Figure 1) shows flight relevant information and a simulated environment as depicted from inside the cockpit looking out of the windscreen (perspective view). The bottom half of the display represents the cockpit panel with indicators for thrust, missile range, target distance, a dynamic navigational display, and various status indicators, such as afterburners and landing gear. In addition, the display contains a dialog box for on-line instructions and feedback that appears on the cockpit panel. The top half of the display shows a simulated world and a head-up display (HUD) containing flight information. In the simulated world, the horizon is depicted as a blue line, and the ground is displayed as a red grid against a black screen simulating night flying. The HUD includes indicators for airspeed (left side), altitude (right side), and heading (bottom) and shows a climb/dive ladder that indicates pitch and roll. Figure 1 also shows how a typical slalom course was depicted.

Tasks

Slalom task. The criterion flight task consisted of an airborne slalom course where participants "flew" the simulator through "gates" in the sky. A gate was represented as four octahedrons arranged in a square suspended in the simulated environment. The gates were positioned so that participants had to turn left or right and climb or dive in order to fly from one gate to the next. Four different courses were created that varied in difficulty. The two easy courses consisted of gates whose centers were 1,000 simulated feet (304.8 m) apart in altitude and averaged 4,000 simulated feet (1,219.2 m) between gates. Given these parameters, participants could generally see the next gate from the gate they were currently flying through. In the two difficult courses, the gates were 2,000 simulated feet (609.6 m) apart in altitude with about 3,200 simulated feet (975.4 m) between gates. These courses required sharper turns and

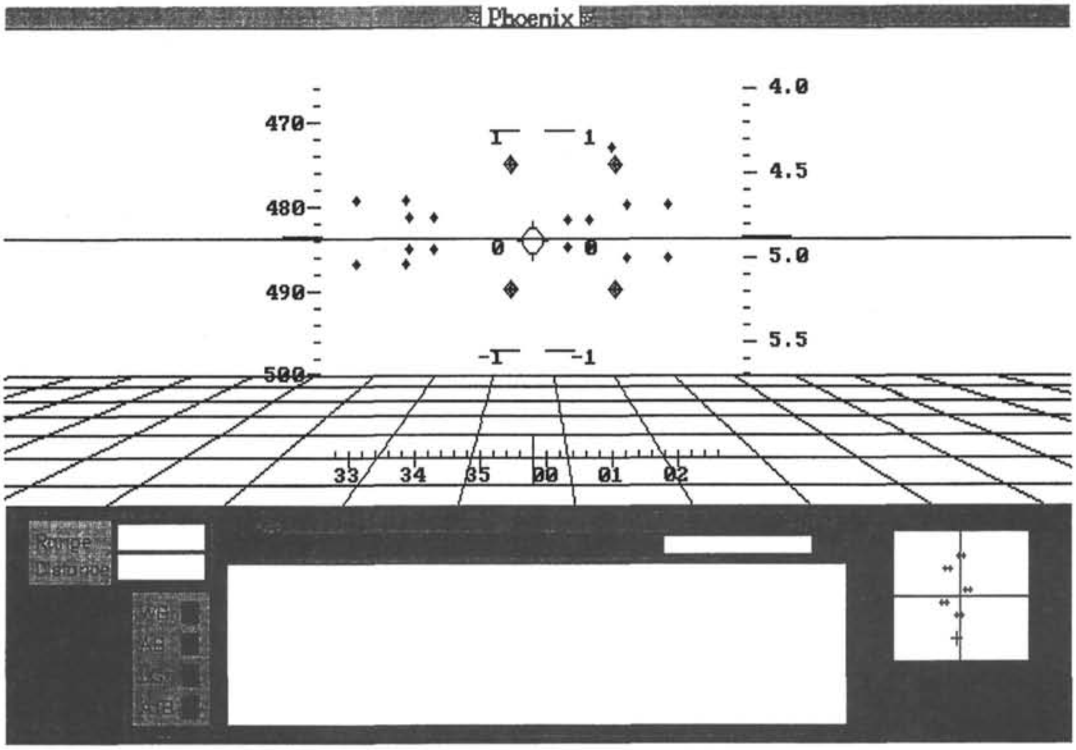

Figure 1. Phoenix computer display.

steeper climbs and dives than the easy courses, and participants rarely saw the next gate from the gate they were flying through. However, for both easy and difficult courses, the radar map showed the horizontal position of the gates at all times. When participants successfully flew the simulator through a gate, the following message was presented in the dialog box at the bottom of the screen: "Good! You made it through gate *n*. Go on to gate *n + 1*." When participants missed a gate, the following message was displayed: "Sorry! You missed gate *n*. Go on to gate *n + 1*." Participants received instructions to proceed to the next gate when they missed one and were not given credit for returning to a missed gate. Only seven gates were displayed at a time. When participants completed the last displayed gate, another set of seven gates was presented and so on until the end of the trial. Each trial lasted 3 mins. Participants received instructions to fly through as many gates as possible while minimizing misses. After each trial, participants received information on how many gates they successfully flew through.

Component tasks. There were a total of 19 different component tasks derived from a systematic analysis of skills involved in the slalom task. These tasks represented a hierarchy of skills ranging from elementary "stick and rudder" skills (e.g., controlling pitch and roll) to more complex ones (e.g., spatial orientation). The following description of component tasks represents the order in which participants performed them.

One of the most basic skills in flying involves maneuvering the plane back to straight and level flight. This skill is required to recover from unusual attitudes (i.e., nonzero pitch and/or roll). Thus, the first three tasks (unpitch, unroll, and unpitch–roll) were intended to train participants on recovering from nonzero pitch and roll. Trials were started with the simulator pitched up or down (unpitch), rolled left or right (unroll), or pitched and rolled (i.e., unpitch–roll). Participants were given 30 s to bring the pitch, roll, or both aspects back to level. Trials were presented in blocks of 10 trials.

The next three tasks represented additional basic skills required for maneuvering the plane (i.e., pitching and rolling skills). In all three tasks, trials started with the simulator flying straight and level. Participants were instructed to pitch and/or roll the plane to designated values. In the pitch task, participants pitched the plane up or down; in the roll task, participants rolled the plane left or right. In the pitch–roll task, participants pitched and rolled the plane to specified degrees. As in the previous tasks, participants had 30 s for each trial, and trials were presented in blocks of 10 trials.

In the next three tasks, participants adjusted their altitude, heading, or both. These tasks are slightly more complex than the previous ones because they require a combination of the aforementioned basic skills. For example, changing altitude requires individuals to pitch the plane up (or down), watch the altimeter while ascending (or descending), and level out at the proper altitude by bringing the pitch back to zero. Thus, skill in changing altitude and heading depended on the basic skills emphasized in the first six tasks. In the current group of component tasks (i.e., altitude, heading, and altitude-heading), the simulator was started in a straight and level orientation, and participants achieved a new altitude, heading, or both. Once attained, the new altitude and/or heading had to be maintained for 5 s. Participants had 2 min for each trial, and trials were presented in 10-trial blocks.

The next six tasks emphasized gate-aiming skills, requiring participants to fly through gates in the sky. Starting from a position in front of a single gate, participants flew the simulator through the gate as quickly as possible. These six tasks represent difficulty manipulations on three dimensions: sharpness of turn, airspeed, and gate size. In the "easy gate" task, the starting positions were almost directly in front of the gate, and participants needed to make only small turns. In the "hard gate" task, participants had to make sharper turns to fly through the gates. In the "slow gate" task, the speed of the simulator was fixed at half thrust; in the more difficult version, "fast gate," the simulator was fixed at full thrust. The "big gate" task consisted of gates that were 50% larger than in the slalom task; in the "tiny gate" task, gates were 50% smaller. These tasks were designed to provide practice on approaching gates from various angles, at various speeds, and with varying degrees of precision. Participants had 30 s to fly through the gates comprising these component tasks.

The final four tasks focused on spatial orientation skills. In these four tasks, participants located a gate on the radar map and flew through it. Successful performance on this task required integration of information from the top-down view of the navigational display with the perspective view of the cockpit windscreen. In these tasks, one gate was located in each quadrant of the airspace. Trials started at random positions in the environment, and participants located their position and the position of the specified gate, and then flew through the gate. In two of the four tasks (Orient-Plan I and Orient-Plan II), participants had unlimited time to plan their course. That is, they were first shown their position and told which gate to fly through. Participants then initiated each trial themselves after they had planned their course to the gate. In the other two tasks (Orient I and Orient II), participants did not have time to plan their course. Rather, they were told which gate to fly through, but were not shown their position until after the trial had started. Thus, once the trial began, participants had to orient themselves in space quickly and find the correct gate. The roman numerals following the task names represent difficulty level, with I indicating easy versions and II indicating difficult versions of the tasks. Difficulty on these trials was defined in terms of the position of the plane relative to the target gate, how easy it was to find the correct gate, and how sharp a turn was needed. In all four tasks, participants had 30 s to fly through the correct gate.

Cognitive ability tests. Three tests were used to assess spatial aptitude from an on-line battery of cognitive ability measures (CAM-4; Kyllonen et al., 1990). These spatial tests assessed participants' working-memory capacity, information-processing speed, and inductive reasoning skill, using spatial (as opposed to verbal or quantitative) stimuli. All tests were presented on the same computers used for the Phoenix task.

The working memory test was a four-term ordering task with blocks as the stimuli. In this test, participants were required to relate what was described in three pictorial statements to the sequence of four block figures. Figures consisted of blocks divided by a diagonal line and colored pink with black or blue with black (e.g., one side of diagonal is pink, and the other is black). The direction of the diagonal could change positions, allowing for different combinations (e.g., a diagonal going from the top left to the bottom right may cause pink to be on the top and black to be on the bottom; a diagonal going from the top right to the bottom left may cause the black to be on top and the pink to be on the bottom). For each statement, two blocks of the same color (i.e., either pink/black or blue/black) appeared with an arrow. The arrow described the sequence in which the two blocks should appear (e.g., one on top of the other). The arrow could have had a slash through it, which was interpreted as meaning *not* (e.g., Block 1 will not appear below Block 2). The third statement merely displayed solid pink and solid blue blocks, describing the sequence of the pink and blue blocks (e.g., pink will not appear before blue).

The pictorial statements appeared one at a time at the top of the screen. Participants determined the sequence of blocks as the statements appeared. After the final statement, eight numbered alternatives appeared on the screen with a timer. These alternatives expressed the possible combinations using the presented statements, and the participants were required to type the number corresponding to the correct sequence. Correct responses were followed by music, incorrect responses caused a buzzer to be sounded, and the three statements were then displayed to show how the item was incorrect. Next, three asterisks appeared to warn participants that the next item was about to be presented. This test contained 24 items. The alternative responses were 1–8 using the number keys at the top of the keyboard.

The information-processing speed test used similar stimuli as the working memory test above, but differed in that it consisted of a two-term ordering test. Participants decided as quickly as possible whether the presented figure combinations matched the simple sequence formula specified by figure statements initially provided on the screen. Shortly after each figure statement was presented, a set of two blocks was shown in the middle of the screen. If the figure and the initial statement matched, *L* was the correct response; if they did not match, *D* was the correct response. Upon a correct response, music was sounded. Incorrect responses were followed by a buzzer. The next item, preceded by three warning asterisks, was then presented. At the end of the test, the percentage correct and average response time were displayed. This test contained 12 items.

In the inductive reasoning test, participants were shown a 3 × 3 matrix in which a figure was contained in all but one of the cells. Participants looked at the figures and applied horizontal and/or vertical rules to determine what figure belonged in the empty cell. The matrix was shown on the screen concurrent with the eight alternative responses. Participants typed in the alternative they believed corresponded to the matrix. Some of the themes used were gradual shading of figures, alternating positions of figures (e.g., a square, circle, and triangle; a circle, triangle, and square; a triangle, square, and circle), successive movements of shapes to form new figures, successive additions or deletions to figures, and rotation of figures. All themes were used in combination to form both horizontal and vertical rules for participants to induce. Participants determined the missing figure as quickly as possible without sacrificing accuracy. After entering a response, participants were informed whether it was correct. If no response was entered, the item was counted wrong. Participants had 10 min to solve all nine problems contained in this test. There were eight alternative responses using the 1–8 keys at the top of the keyboard.

Procedure

Participants completed training and testing in groups of 18 to 22. Approximately half of each group was assigned randomly to each of the two experimental groups. The study took 3 consecutive days to complete. All training was completed on Day 1. Participants signed a consent form, completed the CAM tests, and then received Phoenix instructions and a brief computer-based introduction to the Phoenix simulator. This introduction familiarized participants with the Phoenix displays and controls. All participants then received a pretest consisting of 4 trials on the slalom task (1 trial for each course). This pretest not only served to assess initial ability but also provided participants the criterion task context for subsequent training (Carlson et al., 1990; Singley & Anderson, 1989). Following the pretest, the whole-task group continued to practice on the slalom task for five blocks of 16 trials each. Each block of trials was separated by either a 15-min rest period or a 1-h lunch period.

For the part-task group, the component skills were presented in the order in which they were described above, and all were self-paced. However, total time on training was equated for all participants in the part-task group by requiring faster participants to repeat trials when they finished early. Equating part-task participants on total time, rather than equating total number of trials (or some other objective criterion), poses a potential problem. Skilled participants who completed the task faster would receive more practice trials compared with the less skilled participants. However, this same potential problem existed for the whole-task group in that highly skilled participants would be able to fly faster and complete more gates per trial compared with the lesser skilled participants. Hence, the two groups were treated similarly with regard to practice trials. Moreover, our procedure equated all participants on total time on the simulator and allowed us to synchronize rest and lunch periods between the two groups.

On Day 2, the transfer phase, participants received three blocks of 16 trials on the slalom task. As during training, blocks lasted approximately 50 min and were separated by 15-min breaks. This transfer phase was the first time since the pretest that the part-task group had performed the slalom task. On Day 3, the backward-transfer phase, all participants completed the component skills tasks. Similar to Day 1, the component tasks were presented in five 1-h blocks separated by either a 15-min break or a 1-h lunch period.

Results

Spatial Aptitude

To explore the role of spatial aptitude on training effectiveness, we conducted a factor analysis on the accuracy scores on the three spatial aptitude tasks (i.e., working memory, processing speed, and inductive reasoning). Because of difficulties in the data collection process, only 65 participants had valid data on the spatial aptitude tasks (32 participants in the whole-task group and 33 in the part-task group). With a principal axis factor analysis, one factor was extracted, and the solution accounted for 67% of the variance. Factor scores were saved, and a median split procedure was used to identify high- and low-aptitude individuals.

Slalom Task Performance

The two criterion measures of performance on the slalom task were speed (i.e., the total number of gates flown through) and accuracy (i.e., total gates successfully completed divided by the total number of gates possible). Of the 16 trials per block, 8 were from easy courses (Trials 1, 2, 5, 6, 9, 10, 13, and 14), and 8 were from difficult courses (Trials 3, 4, 7, 8, 11, 12, 15, and 16). For the purpose of reducing within-subjects variability, we averaged performance across blocks. The data in Figures 2 and 3 represent block averages on easy (see Figures 2a and 3a) and difficult courses (see Figures 2b and 3b). Thus, each point in these figures represents average performance on eight trials in each transfer block. An alpha level of .05 was used for all significance tests.

Speed data. Figure 2 shows average training (whole-task group only) and transfer scores on easy (see Figure 2a) and difficult (see Figure 2b) courses for high- and low-aptitude groups. Separate 2 (training condition) × 2 (aptitude) × 3 (block) multivariate analyses of variance (MANOVAs),[2] with block as a repeated measures variable, were performed on easy and difficult trials. For the easy courses, as shown in Figure 2a, whole-task participants performed better (M = 16.41) than part-task participants (M = 12.83), and high-aptitude individuals (M = 15.75) performed better than low-aptitude individuals (M = 13.47). In addition, scores improved from Block 1 (M = 12.07) to Block 2 (M = 15.34) and Block 3 (M = 16.31). The three-way MANOVA supported all these observations by showing significant main effects of training condition, $F(1, 61) = 6.70$, $p = .012$, spatial aptitude, $F(1, 61) = 4.08$, $p = .048$, and block, Wilks's exact $F(2, 60) = 42.52$, $p = .001$. None of the interactions reached statistical significance.

For difficult courses, the analysis revealed a significant main effect of block, Wilks's exact $F(2, 60) = 18.07$, $p = .001$, and a significant Training Condition × Spatial Aptitude interaction, $F(1, 61) = 4.27$, $p = .043$. The main effect of block indicates participants improved across transfer blocks (M = 8.72, 10.69, and 10.56 for Blocks 1, 2, and 3, respectively). Further analysis of the Training Condition × Spatial Aptitude interaction revealed a main effect of training condition for low-aptitude individuals, $F(2, 31) = 5.54$, $p = .025$, but not for high-aptitude participants, $F(2, 30) <$

[2] The MANOVA procedure was used for repeated measures analysis as recommended by O'Brien and Kaiser (1985).

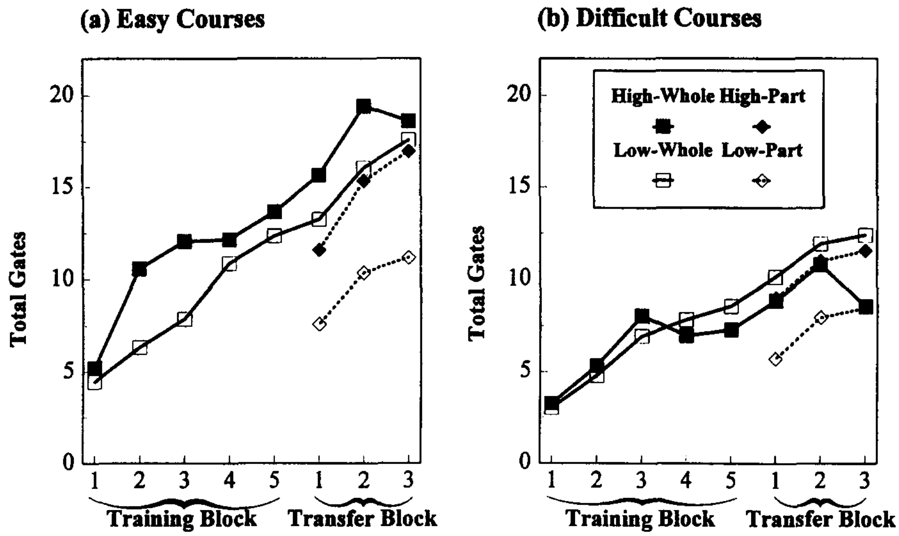

Figure 2. Total number of gates flown through by high- and low-aptitude participants in the whole-task training and part-task training groups on the slalom task.

1.0. These findings and the general pattern in Figure 2b suggest that low-aptitude participants did not perform well following part-task training.

Accuracy data. Figure 3 shows the average accuracy on easy (see Figure 3a) and difficult (see Figure 3b) courses during the transfer phase. In this figure, the abscissa differs from that in Figure 2 because the training data are not shown. These data were submitted to the same 2 × 2 × 3 MANOVAs that were used to analyze the speed data. Data from the easy courses showed that high-aptitude individuals performed more accurately (M = 73.4%) than low-aptitude individuals (M = 56.8%). This difference was supported by a significant main effect of spatial aptitude, $F(1, 61) = 13.78, p = .001$. The analysis also revealed a main effect of block, Wilks's exact $F(2, 60) = 5.74$, $p = .005$, and a Spatial Aptitude × Block interaction, Wilks's exact $F(2, 60) = 3.28, p = .044$. The interaction was further examined by testing the

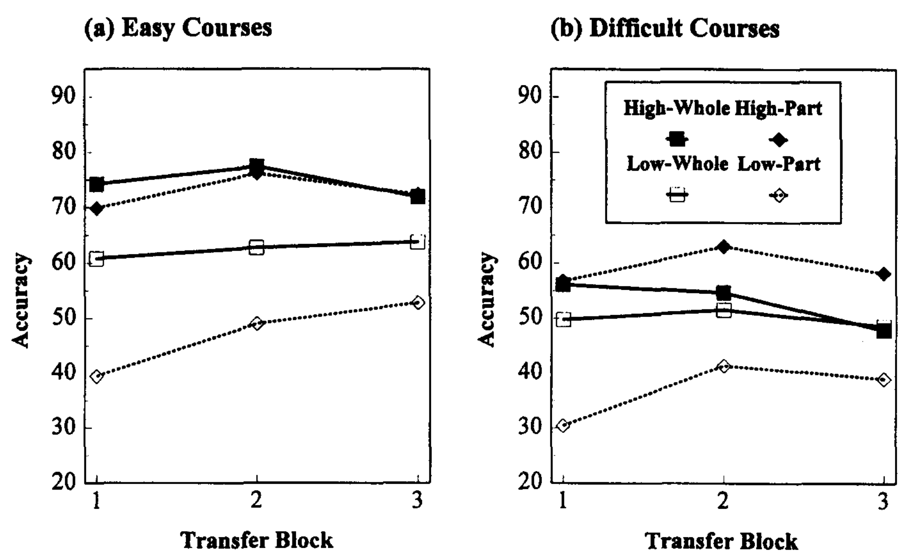

Figure 3. Accuracy performance (i.e., percentage of gates made) of high- and low-aptitude participants in the whole-task training and part-task training groups on the slalom task.

simple main effect of spatial aptitude for each transfer block separately. These analyses revealed that the main effect of spatial aptitude was very strong for Block 1, $F(1, 61) = 14.70, p = .001$, and Block 2, $F(1, 61) = 15.27$, $p = .001$, relative to Block 3, $F(1, 61) = 7.11, p = .011$. These analyses reveal that differences in spatial aptitude were attenuated across transfer blocks but were not eliminated.

For difficult courses, the 2 × 2 × 3 MANOVA revealed significant main effects of spatial aptitude, $F(1, 61) = 7.92, p = .007$, and block, Wilks's exact $F(2, 60) = 5.86$, $p = .005$. In addition, training condition interacted with spatial aptitude, $F(1, 61) = 4.68, p = .034$, and block, Wilks's exact $F(2, 60) = 4.15, p = .020$. To examine the Training Condition × Spatial Aptitude interaction, the main effect of training condition was examined for low- and high-aptitude individuals separately. Low-aptitude individuals showed better performance following whole-task training ($M = 49.9\%$) than part-task training ($M = 36.9\%$), $F(1, 31) = 4.38$, $p = .045$. High-aptitude individuals performed slightly better following part-task training ($M = 59.3\%$) than whole-task training ($M = 52.9\%$), although this difference was not statistically significant, $F(1, 30) < 1.0$.

The Training Condition × Block interaction was examined by testing the simple main effect of block for each group separately. Individuals receiving part-task training showed significant improvement across the three test blocks, Wilks's exact $F(2, 30) = 11.03$, $p = .001$. Whole-task individuals showed a decline in performance across test blocks that approached statistical significance, Wilks's exact $F(2, 30) = 3.20, p = .055$. The improvement in scores by the part-task group may be due in part to the lower overall performance on the initial posttest block. The slight decline in performance by the whole-task group may be due in part to fatigue.

Differential transfer. To determine the amount of transfer, we used the formula of Katona (1940): percentage transfer $= 100 \times (E_t - C_i)/(C_t - C_i)$, where E_t is the average performance of the part-task group on the criterion task during transfer, C_i is the average performance of the whole-task group during initial training, and C_t is average performance of the whole-task group during transfer. For easy courses, $E_t = 10.13$ gates, $C_i = 4.71$ gates, $C_t = 14.07$, and the percentage transfer is 57.9%. For difficult courses, $E_t = 7.78$ gates, $C_i = 5.48$ gates, $C_t = 9.68$, and the percentage transfer is 54.9%.

Component Task Performance

Next, we wished to examine performance on the component tasks. Positive transfer from the whole task to the component tasks indicates which of the component tasks are important to the whole task. That is, if the whole-task group performs better on a component task during transfer than the part-task group does during initial training, then we may infer that the component task involves skills acquired during whole-task training. These skills are presumed to be important for skilled performance on the whole task.

Analysis of component factors. Rather than conduct separate significance tests for each of the 19 component tasks, we wished to reduce family-wise error rate by performing analyses on a small set of composite scores created by combining similar component tasks. One method of generating composite scores would be to combine component tasks that were based on our task hierarchy. Instead, we performed a global factor analysis on component task scores across all participants. Even though a global factor analysis could be criticized for masking important differences between the groups, we used the global analysis for two reasons. First, we wanted to combine component tasks that were based on empirical data rather than theory. Second, factor scores are attractive in that they are standardized scores ($M = 0$, $SD = 1$). Running separate factor analyses for each group makes it impossible to compare groups because the means and standard deviations would be identical across groups. In summary, our goal was not to create definitive factor solutions for each group but to reduce the number of significance tests, the family-wise error rate, and the complexity of the interpretation.

Using the principal-axis factoring procedure with varimax rotation, we extracted five factors that accounted for 80.0% of the variance. The rotated factor matrix is shown in Table 1. As shown in this table, Factor 1 loads strongly on the six gates tasks. Recall that these tasks required individuals to fly through a single gate of variable size from different angles and speeds. Hence, Factor 1 is Gate Aiming. The second factor loads most heavily

Table 1
Task Loadings on the Five Rotated Factors Extracted From the Analysis of the Backward-Transfer Group

Task	Factor: Gate aiming	Spatial orientation	Recovery	Altitude-heading	Roll
Gate					
Hard	.87	.15	.06	.21	.08
Easy	.86	.26	.13	.18	.09
Slow	.85	.27	.16	.12	.17
Big	.80	.25	.26	.06	.15
Tiny	.76	.13	.22	−.14	.16
Fast	.71	.32	.08	.08	.27
Orient					
Plan I	.17	.82	.15	.16	.19
II	.38	.79	.22	.22	.02
I	.34	.73	.21	.22	.17
Plan II	.24	.72	.04	.10	.03
Unpitch	.11	.17	.78	.23	.04
Unroll	.08	−.01	.76	.05	.08
Unpitch–roll	.14	.18	.74	.28	.18
Pitch	.24	.20	.59	.01	.09
Heading	.04	.17	.15	.84	.06
Altitude-Heading	.21	.29	.29	.72	.11
Altitude	.15	.38	.38	.41	.26
Roll	.20	.13	.17	.03	.85
Pitch–roll	.41	.14	.16	.23	.73

on the four spatial orientation tasks. The distinguishing feature of these tasks is that they all require the individual to use the radar map in order to navigate to a specific gate. Thus Factor 2 is Spatial Orientation. Factor 3 loads most strongly on the three tasks requiring participants to "recover" from nonzero pitch and roll to return to straight and level flight (i.e., unpitch, unroll, and unpitch–roll). In addition, this factor also loads on the pitch task. Because of the nature of the first three tasks, this factor is called Recovery. Factor 4 loads most strongly on the two tasks requiring individuals to adjust their heading and altitude (i.e., heading and altitude-heading). For this reason, Factor 4 is Altitude-Heading. The fifth factor loads almost exclusively on the two tasks that require individuals to roll to a specific angle. In these tasks, individuals judge their roll by observing the slant of the horizon line in the simulated environment. There is not a digital indicator for roll, so individuals must be given on-line feedback about their roll. This factor is Roll.

It is interesting to note that the groupings derived from the factor analysis are not very different from what we would have derived from the hierarchy. Had we grouped tasks by the hierarchy, the groupings would have been recovery tasks (unpitch, unroll, and unpitch–roll), pitch–roll tasks (pitch, roll, and pitch–roll), altitude-heading tasks (altitude, heading, and altitude-heading), gates tasks (all six gates tasks), and spatial orientation tasks (all four). Notice that the only difference between this grouping and the one derived from the factor analysis is that the pitch task is grouped with the recovery tasks in the factor analysis solution but is included with the pitch–roll tasks in the hierarchy-derived groupings.

Average factor scores, separated by training condition and spatial aptitude, are shown in Figure 4. Because factor scores are standardized, the data in Figure 4 can be viewed as points above and below the mean in standardized units. The five factors were submitted to a 2 (Training Condition) × 2 (Spatial Aptitude) MANOVA. This analysis revealed significant effects of training condition, Wilks's exact $F(5, 57) = 5.06, p = .001$, spatial aptitude, Wilks's exact $F(5, 57) = 2.79, p = .025$, and the interaction, Wilks's exact $F(5, 57) =$

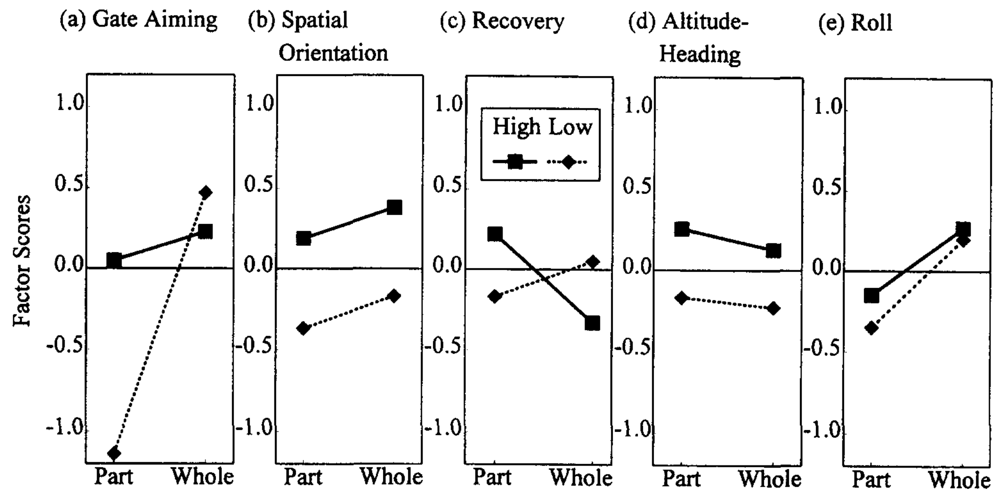

Figure 4. Factor scores of high- and low-aptitude participants in part- and whole-task groups on the five factors extracted from the component tasks.

2.85, $p = .023$. Univariate tests for the main effect of training condition were significant only for the Gate Aiming factor, $F(1, 61) = 18.43, p = .001$, and Roll, $F(1, 61) = 4.17, p = .046$. As shown in Figure 4, the whole-task group performed better than the part-task group on gate-aiming and roll tasks. The univariate tests for the main effect of spatial aptitude were only significant for gate aiming, $F(1, 61) = 5.21, p = .026$, and spatial orientation, $F(1, 61) = 5.44$, $p = .023$. Low-aptitude individuals performed worse than high-aptitude individuals on both components. For the interaction between training condition and spatial aptitude, the only univariate test that reached significance was for gate aiming, $F(1, 61) = 11.84, p = .001$. The data in Figure 4a show the nature of the interaction. Low-spatial ability participants in the part-task group performed much worse on the gate-aiming tasks than any other group. This observation is supported by a significant main effect of training condition for low-aptitude participants, $F(1, 31) = 25.36, p = .001$, but not for high-aptitude participants, $F(1, 30) < 1.0$. The extremely poor performance of the low-aptitude participants in the part-task group may account for the significant main effects of training condition and spatial aptitude observed for gate aiming.

Backward transfer. Although factor scores are attractive for the reasons cited above, they are not feasible for the calculation of transfer. Thus, to estimate the amount of transfer from the whole task to the component task factors, we generated composite scores by averaging across the component scores in each factor derived from the factor analysis. Using these composite scores, we computed with Gagné, Forster, and Crowley's (1948) modification of Katona's (1940) equation: percentage transfer = $100 \times (P_{bt} - P_i)/(T - P_i)$, where P_{bt} is the average performance of the whole-task group on the component tasks during backward transfer, P_i is the average performance of the part-task group during initial training, and T is the total possible score on the component tasks. We modified the notation so that in backward transfer analysis, the part-task group is the control or comparison group instead of the whole-task group, as in the preceding transfer analysis. In addition, we used the total possible score on the component tasks (i.e., 100%) instead of part-task group scores on component tasks during backward transfer because subjects in the part-task group had lengthy experience on the component tasks during training and were thus more likely to approach the maximal possible score on each component.

Table 2 shows the means and percentage transfer scores for the five composite scores. Notice that Gate Aiming and Roll show the largest amount of transfer. This is consistent with the findings that the whole-task group performed significantly better on the gate-aiming and roll tasks during transfer than the part-task group did during training. It is also worth noting that two of the factors (i.e., Recovery and Heading) produced negative transfer. That is, the part-task group performed better

Table 2
Composite Scores and Estimates of Transfer From the Whole Task to the Component Tasks

Factor	Score: Part task	Score: Whole task	Transfer (%)
Gate Aiming	68.57	88.41	63.12
Spatial Orientation	35.12	40.79	8.74
Recovery	83.88	83.10	−4.84
Heading	77.31	76.13	−5.20
Roll	41.67	61.86	34.61

Note. The total possible score for each factor was 100%.

on these factors in training than the whole-task group did during backward transfer.

Discussion

The primary results of this experiment can be summarized as follows: (a) whole-task training was superior to part-task training on speed measures for easy courses, (b) low-aptitude individuals in the part-task training group performed more poorly than low-aptitude individuals in the whole-task group on difficult courses, (c) the 19 different component tasks reduced to five unique factors, and (d) only two of the five factors (Gate Aiming and Roll) appeared to be related to proficiency on the slalom task.

From these results, we have concluded that the part-task training procedure was only moderately effective. That is, the part-task training group showed some transfer from the component tasks to the slalom task, but the amount of transfer was rather modest (55% to 58%). The modest benefits of part-task training observed in the present study are rather typical of much part-task training approaches in the literature (Wightman & Lintern, 1985).

As discussed in the introduction, one explanation for why transfer of the part-task training group was less than expected is that the component tasks did not adequately represent the critical skills underlying the criterion task. Of the five factors we extracted, only two (Gate Aiming and Roll) showed substantial backward transfer from the whole task to the part tasks. This finding suggests that only the Gate Aiming and Roll factors involve skills that are directly related to the criterion task. One factor (Spatial Orientation) showed small, nonsignificant backward transfer, suggesting that the part-task group may have accrued small benefits from these component tasks. Two of the five factors (Recovery and Heading) showed small, though nonsignificant, negative backward transfer, suggesting that presentation of these tasks may have a slightly disruptive effect on learning. Overall, the results suggest that a substantial portion of the component tasks did not produce learning benefits. Elimination of the nonbeneficial component tasks may produce a more efficient part-task training regime.

Results from the gate-aiming analysis (see Figure 4) are particularly interesting with regard to the Training Condition × Spatial Aptitude interaction. Low-aptitude individuals in the part-task group performed dramatically worse than low-aptitude individuals in the whole-task group. High-aptitude individuals showed no effect of training condition. This pattern was also obtained on difficult courses in the criterion task. These findings contradict our prediction that low-aptitude individuals would perform better in part- than whole-task training because they would be overwhelmed by cognitive demands of whole-task training. Rather, these findings suggest that the low-aptitude individuals were particularly sensitive to the inefficiencies in the part-task training condition as revealed by the backward-transfer analysis. Stated another way, high-aptitude individuals were able to overcome the inefficiency of part-task training, but low-aptitude individuals were not. Thus, improving the efficiency of part-task training should have a disproportionately greater benefit for low-aptitude individuals than high-aptitude individuals. In other words, we can specifically posit an Aptitude × Treatment interaction (Cronbach & Snow, 1977; Shute, 1992).

Another interesting finding that emerged was the failure of spatial orientation to show a strong effect of training condition. Because high-aptitude individuals performed better than low-aptitude individuals on both the criterion task and the spatial orientation tasks, it is tempting to conclude that spatial orientation is important for the criterion task. However, there was no effect of training condition on the spatial orientation tasks. This suggests that practice on the criterion task did not improve performance on the spatial orientation tasks. One possible interpretation of these findings is that spatial orientation skill is not critical to the slalom task. Note that this conclusion, however, is only correlational. An alternative interpretation is that spatial orientation is critical, but not malleable in the context of this study (i.e., it represents a fixed ability rather than one that can be manipulated through instructional environment).

Experiment 2

In general, the results from the component task analysis from Experiment 1 suggest that a significant portion of the part-task training regime was devoted to tasks that were not critical to the criterion task. Of the five factors identified in the factor analysis, only Gate Aiming and Roll were strongly related to proficiency on the slalom task. Moreover, two factors (Recovery and Altitude-Heading) appeared to show negative backward transfer, suggesting that part-task training on them may have disrupted learning on the criterion task. These findings led us to hypothesize that by eliminating the "deadwood" tasks, the effectiveness of the part-task training regime would be improved. The goal of Experiment 2 was to provide a direct test of this hypothesis.

Specifically, Experiment 2 focused on two hypotheses suggested by Experiment 1. The first hypothesis is that a part-task training program that concentrates on gate-aiming skills will be more effective than one that does not. Second, individuals showing lower proficiency on the gate-aiming skill should show larger benefits from concentrated practice compared with participants with higher proficiency. To test these hypotheses, we compared a part-task training program, focusing on the gate-aiming component tasks, with one that included altitude-heading and spatial orientation tasks. Both of these part-task conditions were compared with a whole-task training condition.

Method

Participants and Equipment

Participants consisted of 66 men and 66 women recruited by local temporary employment agencies in San Antonio, Texas. Participants were paid about $5.00 an hour. They ranged in age from 18 to 30 years of age and reported spending less than 20 hr per week playing video games (*Mdn* = 1.25 hr). Of the 132 participants, 13 did not complete the study. Experiment 2 used the same hardware and software as Experiment 1.

Tasks

The slalom task was identical to that used in Experiment 1. In addition, 12 of the 19 component skills tasks used in Experiment 1 were selected (on the basis of the factor analysis) and used in the part-task training conditions. One part-task training group received practice on the six gate-aiming tasks. In these tasks, participants flew through a single gate from various angles and speeds. Because the gate-aiming tasks could be completed more quickly than the altitude-heading and spatial orientation tasks, a seventh task (i.e., the pitch task from Experiment 1) was included in the gate-aiming regime to equate the total training time. The second part-task training group received practice on the three tasks comprising the altitude-heading factor (i.e., Altitude, Heading, and Altitude-Heading) and two of the four tasks from the Spatial Orientation factor (i.e., locate a specific gate on the radar map and then fly through the gate).

Procedure

Participants, randomly assigned to one of three groups, (a) whole-task training, (b) part-task training on the gate-aiming tasks, and (c) part-task training on altitude-heading tasks, took part in a 2-day study. Individual attrition resulted in slightly unequal groups. The group receiving whole-task training consisted of 20 men and 20 women, the group receiving gate-aiming part-task training had 18 men and 19 women, and the group receiving altitude-heading part-task training had 21 men and 21 women.

On Day 1, participants signed a consent form and then proceeded to take the spatial ability pretests. After completing these tests, participants received general instructions on the Phoenix simulator, a pretest on general Phoenix skills, and a pretest on the slalom task. The slalom pretest provides important contextual information for subsequent training (Anderson, 1983; Carlson et al., 1990). The general flight skill pretest consisted of one trial in which the participants changed the thrust of the aircraft, two trials in which they changed pitch, and two in which they changed roll. The pretest on the slalom task consisted of two trials on the easy courses and two trials on the hard courses. Following the pretests, participants received three 1-hr blocks of training in their respective regimes, each separated by a 15-min break. On Day 2, participants completed the transfer phase on the slalom task. Transfer consisted of four 30-min blocks of trials on the slalom task with easy and difficult courses.

Results

Spatial Aptitude

To identify and group participants into high- and low-spatial ability groups, we conducted a factor analysis on the accuracy scores on the three spatial aptitude tasks (i.e., inductive reasoning, processing speed, and working memory). Because of difficulties in the data collection process, only 88 participants had valid data on the spatial aptitude tasks. With a principal axis factor analysis, one factor was extracted, and the solution accounted for 58.2% of the variance. Factor scores were saved, and a median split procedure was used to identify high- and low-aptitude individuals. This procedure resulted in having unequal numbers of participants in each group. The whole-task group consisted of 13 low-aptitude individuals and 18 high-aptitude individuals. The gate-aiming group had 18 low- and 16 high-aptitude participants. The altitude-heading group had 13 low- and 10 high-aptitude individuals.

Pretests

Nine pretest measures were assessed. Five of the pretest measures came from the general Phoenix skills pretest: one thrust task latency, two pitch latencies, and two roll latencies. The other four pretest measures came from the slalom pretest: two slalom speed scores (one mean score for easy courses and one mean score for hard courses) and two slalom accuracy scores (mean scores for the easy and difficult courses, respectively). To reduce the number of pretest measures, we performed a principal axis factor analysis on the nine pretest measures. Three factors were extracted and rotated with the varimax rotation procedure. The solution accounted for 75.6% of the variance. The first factor, Slalom Pretest, consisted of the four slalom task measures, which were weighted positively and strongly (>.85). The second factor, Pitch Pretest, consisted of the two pitch pretests and the thrust pretest, with most of the weight on the down-pitch pretest (.85), followed by the up-pitch pretest (.53), and the thrust pretest (.46). The third factor, Roll Pretest, consisted of the two roll pretests with the right roll pretest showing a larger weight (.74) than left roll pretest (.49). Factor scores were saved for use as covariates in subsequent analyses.

Although multivariate and univariate tests showed that groups did not differ on the pretest factor scores, we tested whether the pretest scores could be used as covariates. We first calculated the correlations between the pretest factor scores and the measures of performance on the criterion tasks. The first two factors (i.e., Slalom Pretest and Pitch Pretest) showed significant correlations with all the performance measures, but the third factor (i.e., involving roll) was not significantly correlated with any of the performance measures. Next, we tested the assumption of equivalent slopes and concluded that the first two factors did not violate this assumption. Therefore, in subsequent analyses of slalom task performance, we included Slalom Pretest and Pitch Pretest scores as covariates.

Component Task Performance

Performance during part-task training was about equal for the two part-task training groups. The gate-aiming group showed slightly higher accuracy during training (M = 81.4%, SD = 17.1) than the altitude-heading group (M = 75.2%, SD = 18.1). However, because the two groups performed different tasks, direct statistical comparisons between the two groups cannot be made.

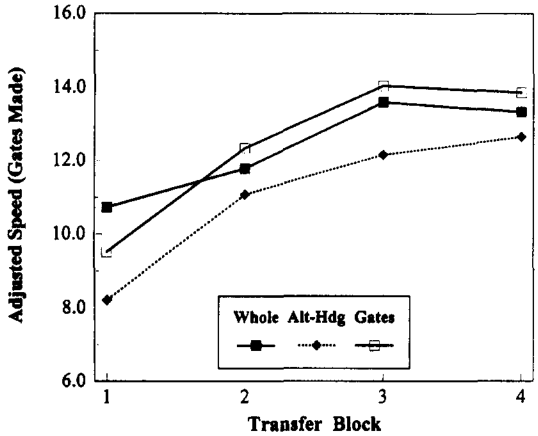

Figure 5. Speed scores of high- and low-aptitude participants across four transfer blocks (adjusted by covariates). Alt-Hdg = altitude-heading.

Slalom Task Performance

Figure 5 shows speed measures (i.e., mean number of gates made) on the easy courses on the four transfer blocks in the posttest, adjusted for the covariates. These data were analyzed in a 3 (Training Condition) × 2 (Spatial Aptitude) × 4 (Block) mixed factors MANOVA, with both pretest factor scores as covariates. Block was the only repeated measure. This analysis revealed only a main effect of block, Wilks's exact $F(3, 80) = 24.16$, $p = .001$. Planned polynomial contrasts showed significant linear, $F(1, 80) = 48.52, p = .001$, and quadratic effects, $F(1, 80) = 15.85, p = .001$. For difficult courses, the same pattern was observed, with only a main effect of block reaching statistical significance, Wilks's exact $F(3, 80) = 6.24$, $p = .001$. Planned polynomial contrasts indicated that only the linear effect was significant, $F(1, 80) = 15.88, p = .001$.

Slalom task accuracy scores for the easy courses were adjusted by the pretest factor scores and are shown in Figure 6. Figure 6a displays data for high-aptitude participants; Figure 6b shows data for low-aptitude participants. The 3 × 2 × 4 mixed factors MANOVA revealed main effects of training condition, $F(2, 80) = 3.45$, $p = .037$, spatial aptitude, $F(1, 80) = 8.67$, $p = .004$, and block, Wilks's exact $F(3, 80) = 10.83$, $p = .001$. In addition, the Training Condition × Block interaction was statistically significant, Wilks's exact $F(6, 160) = 3.29, p = .004$. The main effect of spatial aptitude indicates that high-aptitude individuals performed better ($M = 69.6\%$) than low-aptitude individuals ($M = 57.8\%$). The planned comparisons of the main effect of training condition indicate that the whole-task group did not differ from the average of the two part-task groups, $F(1, 80) = 1.08$, $p = .303$, but the gate-aiming group performed better than the altitude-heading group, $F(1, 80) = 6.51, p = .013$.

The main effect of training condition must be interpreted in the context of the training condition by block interaction. This interaction was examined by testing the simple main effect of training condition for each block separately. For transfer Block 1, the main effect of training condition was significant, $F(2, 80) = 5.14, p = .008$, and planned contrasts indicated that the whole-task group performed better than the two part-task groups, $F(1, 80) = 8.54, p = .005$, but the two part-task groups did not differ from each other, $F(1, 80) = 3.20, p = .077$. For Block 2, the main effect of training condition was significant, $F(2, 80) = 4.15, p = .019$, and planned contrasts revealed that the whole-task group did not differ from the average of the two part-task groups, $F(1, 80) < 1.0$, but that the gate-aiming group performed significantly better than the altitude-heading group, $F(1, 80) = 8.29$, $p = .005$. The main effect of training condition approached significance for block 3, $F(2, 80) = 2.95, p = .058$, and was not significant for Block 4, $F(2, 80) = 1.98, p = .144$. Despite the apparent differences between the high- (see Figure 6a) and low-aptitude (see Figure 6b) learners, the interaction between spatial aptitude and training condition was not statistically significant, $F(2, 80) < 1.0$.

For difficult courses, accuracy scores only revealed a main effect of spatial aptitude, $F(1, 80) = 5.44$, $p = .022$. High-aptitude participants performed better ($M = 59.2\%$) than the low-aptitude participants ($M = 50.7\%$). No other main effects or interactions were significant for the difficult courses.

Transfer

To estimate percentage transfer, we again used the formula of Katona (1940): percentage transfer = $100 \times (E_t - C_i)/(C_t - C_i)$. Scores and computed transfer values for the gate-aiming and altitude-heading groups on easy courses are shown in Table 3. Separate transfer scores were calculated for transfer Blocks 1 and 2. In addition,

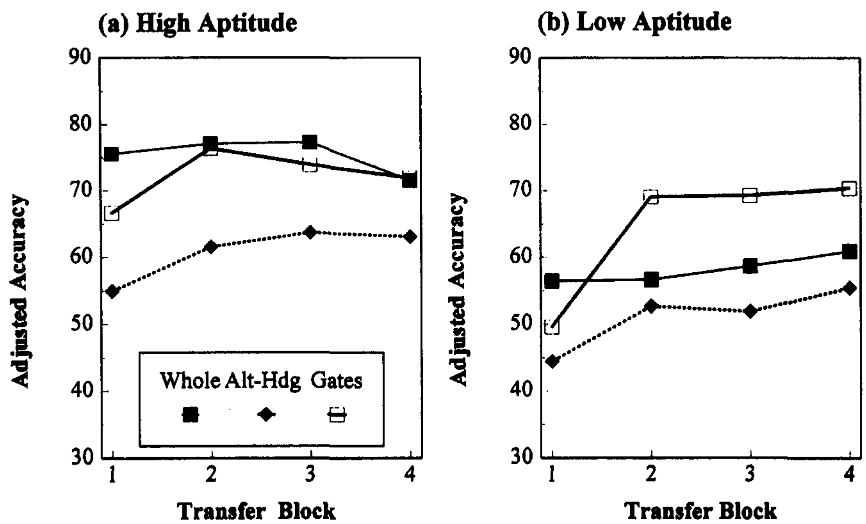

Figure 6. Accuracy scores of high- and low-aptitude participants on four transfer blocks (adjusted by covariates). Alt-Hdg = altitude-heading.

transfer was calculated separately for speed and accuracy data. As indicated in this table, transfer was higher for the gate-aiming group than for the altitude-heading group.

Discussion

The results of this experiment confirm the hypothesis that concentrated practice on the gates tasks leads to better performance on the slalom task compared with practice that focuses on altitude-heading tasks. Overall, the gate-aiming group performed more accurately on the slalom task compared with the altitude-heading group. Additionally, the gate-aiming group scored as well as the whole-task group on the slalom task. Spatial aptitude showed only a main effect on slalom task performance and did not interact with training condition.

It is important to note the conditions under

Table 3
Adjusted Criterion Task Scores and Estimates of Transfer on Each Half of Posttest Block 1 for Speed and Accuracy Data on Easy Courses

Task	Whole-task training	Transfer: Part task	Transfer: Whole task	Transfer (%)
		Speed		
Block 1				
Gate Aiming	5.31	9.57	10.69	79.18
Altitude-Heading	5.31	8.20	10.69	53.72
Block 2				
Gate Aiming	5.31	12.39	11.65	111.67
Altitude-Heading	5.31	11.17	11.65	92.43
		Accuracy		
Block 1				
Gate Aiming	38.66	58.36	65.68	72.91
Altitude-Heading	38.66	49.76	65.68	41.08
Block 2				
Gate Aiming	38.66	72.87	66.75	121.79
Altitude-Heading	38.66	57.04	66.75	65.43

which we observed the benefits of the gate-aiming training. First, the benefits were primarily centered on accuracy. Speed showed a similar pattern but did not reach statistical significance. Second, the benefits of gate aiming were not evident on initial posttest scores. The whole-task group performed better than the part-task groups on the first transfer block. The benefits of gate aiming peaked on the second transfer block. This suggests that gate-aiming training prepared learners to acquire the criterion task more readily than did altitude-heading training. Third, the relative benefits of gate-aiming training waned as the altitude-heading group continued to improve. This is to be expected because transfer trials serve as additional practice trials. Finally, the data in Figure 6 suggest that gate-aiming training was particularly beneficial for low-aptitude learners. They showed nearly the same level of transfer performance as their high-aptitude counterparts, and low-aptitude learners in the other two conditions showed lower performance than their high-aptitude counterparts. However, because the interaction was not significant, we must conclude that the benefits of gate-aiming training did not differ for low- and high-aptitude learners. This contradiction to our prediction will be discussed in conjunction with Experiment 1 findings below.

Perhaps the most impressive finding is the transfer effect. The gate-aiming condition resulted in transfer that exceeded 100% for both speed and accuracy in the second transfer block. The altitude-heading condition produced moderate transfer for speed but not accuracy. Even though direct comparisons of transfer in the two experiments should be made with caution, we observed that the transfer of the gate-aiming group in Experiment 2 (79.1% and 111.6% for speed in transfer Blocks 1 and 2) was greater than that of the part-task group in Experiment 1 (57.9% for easy courses in Block 1). Note that the transfer blocks in Experiment 1 were 60 min in duration and those in Experiment 2 were only 30 min in duration.

Summary and Conclusions

Even though there was an initial advantage for the whole-task group in Experiment 2, it may be concluded that removing the irrelevant component tasks identified in Experiment 1 resulted in a more efficient part-task training regime. Furthermore, the initial advantage of whole-task training shown in transfer Block 1 (Experiment 2) was eliminated relatively quickly. Although statistical comparisons showed there was no difference between the whole-task group and the mean of the two part-task groups, gate-aiming learners showed higher mean scores than whole-task learners on speed and accuracy. This is reflected in the differential transfer scores that exceed 100% in Experiment 2. In Experiment 1, part-task training scores approached but did not exceed, whole-task training scores.

The effectiveness of the gate-aiming regime in Experiment 2 testifies to the value of the backward-transfer procedure in identifying the component tasks that are most critical to the whole task. By showing that only a portion of the tasks in the original part-task regime produced positive backward transfer, we were able to eliminate irrelevant tasks from the part-task training regime. By focusing on the critical tasks, we created a more efficient part-task training regime. We should also point out that we did not select all the component tasks that showed positive backward transfer. The two component tasks comprising the Roll factor in Experiment 1 were not included in gate-aiming training. Also, the four spatial orientation tasks that showed positive, albeit nonsignificant backward transfer, were included in the altitude-heading regime instead of the gate-aiming one. It is interesting to speculate whether these tasks would further improve the efficiency of the gate-aiming condition, but we leave that for future investigation.

Wightman and Lintern (1985) observed that part-task training based on segmentation is more effective than part-task training based on fractionation. Because they were unable to ascertain whether they were observing a general benefit of segmentation or a unique effect of backward chaining, it would be interesting to compare directly the effectiveness of the gate-aiming tasks with that of the Roll factor tasks. Such a comparison would test the question of whether segmentation (i.e., gate-aiming) is more effective than fractionation (i.e., roll). However, the results may be biased in favor of the gate-aiming tasks because they show greater backward transfer than the roll tasks in Experiment 1. To test the question of fractionation versus segmentation, relevant tasks that are equivalent in terms of backward transfer should be selected and then compared directly.

We predicted that manipulating the efficiency of the part-task training condition would have a greater impact on individuals with low- rather than high-spatial aptitude. This prediction was based on the findings of Wightman and Sistrunk (1987). They found that high-aptitude learners are affected less adversely by whole-task training than are low-aptitude learners. Furthermore, this represents a common finding in the ATI literature. That is, ATIs often take the form of large treatment effects for low-aptitude learners and small (or nonsignificant) treatment effects for high-aptitude learners, primarily because of ceiling effects within the high-aptitude population (e.g., Shute, 1992, 1995; Snow, 1994; Swanson, 1990; Tobias, 1994).

Data from the present experiments were partially consistent with these findings. Although we did not find an interaction between training condition and spatial-aptitude as predicted in Experiment 2, we did eliminate the Aptitude × Treatment interaction obtained in Experiment 1. Specifically, in Experiment 1, low-aptitude, but not high-aptitude participants showed a disadvantage for part-task training. We concluded that low-aptitude participants were more sensitive to the inefficiencies of part-task training revealed in the backward-transfer analysis. In Experiment 2, the ATI was not statistically significant, and the gate-aiming group overall did not differ in performance compared with the whole-task group. Thus, although low-aptitude participants did not show significantly greater benefits from gate-aiming training compared with high-aptitude participants, they were able to overcome the disadvantage shown in Experiment 1. In other words, these findings indicate that irrelevant part tasks in a part-task regime can have a negative effect on low-aptitude learners. Thus, if the reason for the introduction of part-task training in the first place was to help low-aptitude trainees cope with the high demands of the whole task, then the present outcome indicates that unless these tasks provide a relevant context for the performance of the whole task, part training may worsen their situation. Such a finding highlights the importance of identifying and removing irrelevant part tasks from part-task training, and the important role that backward transfer can play in creating a training environment in which all learners may excel.

It should be noted that although the altitude-heading condition was worse than the gate-aiming condition, participants in the former group did acquire skills relevant to the criterion task. Moreover, transfer data suggests that the altitude-heading condition produced as much or more transfer (53.6% and 92.4%, for the first and second transfer blocks) as the part-task training regime in Experiment 1 (57.9%). This can be explained by pointing out that the spatial orientation tasks included in the altitude-heading condition showed positive backward transfer to the criterion task in Experiment 1. In other words, the altitude-heading condition included tasks that showed positive transfer as well as negative transfer. There may have been enough relevant tasks to produce as much transfer as in Experiment 1. An alternative explanation is that participants may acquire general abilities about display and control dynamics whenever they "fly" the simulator. These skills may not be directly relevant to the criterion task, but they still may provide the general skills that transfer to criterion performance. However, these arguments do not detract from the conclusion that removing many of the less relevant tasks improves efficiency of part-task training.

One question that can be raised is whether backward transfer is the only way to identify the critical component skills or if some general heuristic can be derived from the present data. Consider, for instance, spatial aptitudes. In both experiments, we found that gate-aiming skill was best predicted by spatial processing speed. In contrast, altitude-heading skill was predicted best by spatial inductive reasoning. Although these findings suggest that spatial processing speed may serve as a predictor of critical skills, it should be noted that the results may not generalize beyond the criterion task of the current study. There are likely to be other flight tasks that are best predicted by spatial inductive reasoning or spatial working memory.

In conclusion, we suggest that the superiority of whole-task training in Experiment 1 may have been the result of including many noncritical component tasks in the part-task training regime of Experiment 1. Experiment 2 supported this conclusion, whereby concentrated practice on critical skills resulted in performance that was equivalent to the whole-task group and produced transfer that was greater than 100%. Moreover, the gate-aiming condition effectively eliminated an ATI obtained in Experiment 1. This suggests that low-aptitude participants benefited more from the revised part-task training regime than did high-aptitude participants. Finally, the present data confirm the utility of using the backward transfer technique to identify critical components in a part-task training regime. We do not intend to suggest that backward transfer alone is the only way to identify critical component skills. We argue that backward transfer should be used in conjunction with other techniques such as cognitive task analysis. Although the cognitive task analysis may provide theoretically driven predictions about the critical tasks, the backward transfer can provide empirical evidence for supporting or refining predictions. Moreover, we contend that any task could benefit from part-task training provided: (a) an appropriate task decomposition is developed and validated, (b) the ordering of critical skills for the part-task regime is established based on findings in the literature, (c) individual characteristics of learners are taken into consideration in developing the training regime, and (d) there is sufficient and effective reintegration of the part tasks into a unified whole task.

References

Ackerman, P. L. (1992). Predicting individual differences in complex skill acquisition: Dynamics of ability determinants. *Journal of Applied Psychology, 77,* 598–614.

Adams, J. A. (1960). Part trainers. In G. Finch (Ed.), *Educational and training media: A symposium of the National Research Council* (pp. 129–149; Publication No. 789). Washington, DC: National Academy of Sciences.

Anderson, J. R. (1983). *The architecture of cognition.* Cambridge, MA: Harvard University Press.

Ash, D. W., & Holding, D. H. (1990). Backward versus forward chaining in the acquisition of a keyboard skill. *Human Factors, 32,* 139–146.

Bailey, J. S., Hughes, R. G., & Jones, W. E. (1980). *Application of backward chaining to air-to-surface weapons delivery training* (AFHRL-TR-79-63). Williams AFB, AZ: Operations Training Division, Human Resources Laboratory.

Briggs, G. E., & Naylor, J. C. (1962). The relative efficiency of several training methods as a function of transfer task complexity. *Journal of Experimental Psychology, 64,* 505–512.

Briggs, G. E., & Waters, L. K. (1958). Training and transfer as a function of component interaction. *Journal of Experimental Psychology, 56,* 492–500.

Carlson, R. A., Khoo, H., & Elliot, R. G., Jr. (1990). Component practice and exposure to a problem-solving context. *Human Factors, 32,* 267–286.

Carlson, R. A., Sullivan, M. A., & Schneider, W. (1989). Component fluency in a problem-solving context. *Human Factors, 31,* 489–502.

Cronbach, L. J., & Snow, R. E. (1977). *Aptitudes and instructional methods.* New York: Irvington.

Fabiani, M., Buckley, J., Gratton, G., Coles, M. G. H., Donchin, E., & Logie, R. (1989). The training of complex tasks. *Acta Psychologica, 71,* 259–299.

Frederiksen, J. R., & White, B. Y. (1989). An approach to training based upon principled task decomposition. *Acta Psychologica, 71,* 89–146.

Gagné, R. M., Forster, H., & Crowley, M. E. (1948). The measurement of transfer of training. *Psychological Bulletin, 45,* 97–130.

Gopher, D., Weil, M., & Siegel, D. (1989). Practice under changing priorities: An approach to the training of complex skills. *Acta Psychologica, 71,* 147–177.

Kanfer, R., & Ackerman, P. L. (1989). Motivation and cognitive abilities: An integrative/aptitude-treatment interaction approach to skill acquisition [Monograph]. *Journal of Applied Psychology, 74,* 657–690.

Kanfer, R., & Ackerman, P. L. (1996). A self-regulatory skills perspective to reducing cognitive interference. In I. G. Sarason, B. R. Sarason, & G. R. Pierce (Eds.), *Cognitive interference: Theories, methods, and findings* (pp. 153–171). Hillsdale, NJ: Erlbaum.

Katona, G. (1940). *Organizing and memorizing.* New York: Columbia University Press.

Kyllonen, P. C., Woltz, D. J., Christal, R. E., Tirre, W. C., Shute, V. J., & Chaiken, S. (1990). *CAM-4: Computerized battery of cognitive ability tests* [Unpublished computer program]. Brooks Air Force Base, TX.

Linn, M. C., & Petersen, A. C. (1985). Emergence and characterization of sex-differences in spatial ability: A meta-analysis. *Child Development, 56,* 1479–1498.

Lintern, G., & Gopher, D. (1980). Adaptive training of perceptual-motor skills: Issues, results, and future directions. *International Journal of Man–Machine studies, 10,* 521–551.

Lohman, D. F. (1979). *Spatial ability: A review and reanalysis of the correlational literature* (Tech. Rep. No. 8). Stanford, CA: Aptitude Research Project, School of Education, Stanford University.

Maccoby, E. E., & Jacklin, C. N. (1974). *The psychology of sex differences.* Stanford, CA: Stanford University Press.

Mané, A. M., Adams, J. A., & Donchin, E. (1989).

Adaptive and part-whole training in the acquisition of a complex perceptual-motor skill. *Acta Psychologica, 71,* 179–196.

Mané, A. M., Coles, M. G. H., Wickens, C. D., & Donchin, E. (1983). The use of additive factors methodology in the analysis of a complex skill. *Proceedings of the 27th Annual Meeting of the Human Factors Society* (pp. 407–411). Santa Monica, CA: Human Factors Society.

McGee, M. G. (1979). Human spatial abilities: Psychometric studies and environmental, generic, hormonal, and neurological influences. *Psychological Bulletin, 86,* 889–918.

O'Brien, R. G., & Kaiser, M. K. (1985). MANOVA method for analyzing repeated measures designs: An extensive primer. *Psychological Bulletin, 97,* 316–333.

Salthouse, T. A., & Prill, K. (1983). Analysis of a perceptual skill. *Journal of Experimental Psychology: Human Perception and Performance, 9,* 607–621.

Schneider, W., & Detweiler, M. (1988). The role of practice in dual-task performance: Toward workload modeling in a connectionist/control architecture. *Human Factors, 30,* 539–566.

Shute, V. J. (1992). Aptitude-treatment interactions and cognitive skill diagnosis. In J. W. Regian & V. J. Shute (Eds.), *Cognitive approaches to automated instruction* (pp. 15–47). Hillsdale, NJ: Erlbaum.

Shute, V. J. (1993). A macroadaptive approach to tutoring. *Journal of Artificial Intelligence in Education, 4*(1), 61–93.

Shute, V. J. (in press). SMART: Student Modeling Approach for Responsive Tutoring. *User Modeling and User-Adapted Interaction.*

Singley, M. K., & Anderson, J. R. (1989). *The transfer of cognitive skill.* Cambridge, MA: Harvard University Press.

Snow, R. E. (1994). Abilities in academic tasks. In R. J. Sternberg & R. K. Wagner (Eds.), *Mind in context: Interactionist perspectives on human intelligence* (pp. 3–37). New York: Cambridge University Press.

Stammers, R. B. (1980). Part and whole practice for a tracking task: Effects of task variables and amount of practice. *Perceptual and Motor Skills, 50,* 203–210.

Swanson, J. H. (1990, April). *The effectiveness of tutorial strategies: An experimental evaluation.* Paper presented the American Educational Research Association, Boston, MA.

Tobias, S. (1994). Interest, prior knowledge, and learning. *Review of Educational Research, 64*(1), 37–54.

Voyer, D., Voyer, S. & Bryden, M. P. (1995). Magnitude of sex differences in spatial abilities: A meta-analysis and consideration of critical variables. *Psychological Bulletin, 117,* 250–270.

Wightman, D. C., & Lintern, G. (1985). Part-task training for tracking and manual control. *Human Factors, 27,* 267–283.

Wightman, D. C., & Sistrunk, F. (1987). Part-task training strategies in simulated carrier landing final-approach training. *Human Factors, 29,* 245–254.

Received March 29, 1995
Revision received January 25, 1996
Accepted February 13, 1996 ■