Journal of Experimental Psychology: Applied
2002, Vol. 8, No. 3, 194–208

1076-898X/02/$5.00 DOI: 10.1037//1076-898X.8.3.194

Ability and Task Constraint Determinants of Complex Task Performance

Phillip L. Ackerman
Georgia Institute of Technology

Anna T. Cianciolo
Yale University

Previous research on basic information-processing tasks has suggested that there may be a dissociation between the underlying process determinants of task performance and associations with ability measures. This study investigates this dissociation in the context of a complex skill-learning task—an air traffic control simulation called TRACON. A battery of spatial, numerical, and perceptual speed ability tests were administered, along with extensive task practice. After practice, manipulations of task requirements and system consistency were introduced. Ability correlations with performance revealed a dissociation between some manipulations that have effects on performance means and the corresponding correlations with reference abilities. Implications for integrating experimental and differential approaches to explaining performance and possible avenues for improved selection measures are discussed.

Ever since the first modern investigations of the ability correlates of information-processing tasks (e.g., Hunt, Frost, & Lunneborg, 1973), the pattern in the literature has been one of high hopes but generally disappointing results. The high hopes have been of the form that it might be possible to explain individual differences in broad intellectual abilities through a parametric examination of individual differences on basic information processes. The source of the generally disappointing results is that, with the exception of two rather controversial programs of research (i.e., the "inspection time" paradigm, see Deary & Stough, 1996, for a review; and slopes calculated from Hick paradigm choice-reaction time tests, see Jensen, 1998, for a review), there have been very few, if any, replicated findings of substantial correlations between simple tasks, such as the S. Sternberg memory scanning task, the Posner letter-matching task, and so on (for a review, see Lachman, Lachman, & Butterfield, 1979), and broad intellectual abilities. The vast majority of what Carroll (1980) called "elementary cognitive tasks" show modest correlations with broad ability factors, and only show more substantial correlations with lower order, or specific factors such as perceptual speed or processing speed (e.g., see Carroll, 1993, for an extensive review). That is, there has been little evidence for an isomorphic relationship between an information-processing decomposition of tasks and correlations between task performance and particular cognitive–intellectual abilities.

Perhaps the most extensive parametric investigation of these issues was conducted by Kyllonen (1985). In a series of studies with several hundred participants, Kyllonen used an additive-factors logic to build a series of information-processing tasks—from simple reaction time to such tasks as the sentence verification task and other tasks that would now be considered to assess aspects of working memory. In general, what was found was that as task complexity increased, correlations with broad abilities increased. In other words, as the information-processing tasks more closely resembled the kinds of tests that are common to ability batteries, the correlations between tasks and broad cognitive abilities similarly increased. Kyllonen's approach was largely a bottom-up approach, starting with the most basic tasks and building in complexity to ability test type tasks.

A similar conclusion can be derived from a top-down approach, such as that by R. J. Sternberg (1977). Sternberg started with an extant ability test domain—that of analogical reasoning—and separated it into components in a parametric fashion. As the task was separated into components (of encoding, mapping, justification, etc.), in general, the correlations between components and overall ability declined, with perhaps one or two exceptions (such as "study time," which was found to have a substantial correlation with overall performance). Nonetheless, it is clear that none of the individual components of analogical reasoning was sufficiently valid to be used separately as an adequate representation of the overall ability level.

Researchers have suggested (e.g., Ackerman, 2000) that the "content" of a task is a much stronger determining factor for ability–performance relations than the underlying processes. Generally speaking, if a task depends on the processing of verbal content, it is likely to have higher correlations with verbal abilities than it does with either numerical or spatial abilities, or with general reasoning ability, even when the underlying process demands are substantially different (e.g., see Ackerman, 1986; Ackerman & Woltz, 1994). Even psychological tests that are designed

---

Editor's Note. Dan J. Woltz served as action editor for this article.—PLA

Phillip L. Ackerman, School of Psychology, Georgia Institute of Technology; Anna T. Cianciolo, Psychology of Abilities, Competencies, and Expertise Center, Yale University.

This research was supported by Air Force Office of Scientific Research Grants F49620-98-1-0031 and F49620-01-0170, Phillip L. Ackerman, principal investigator. Portions of this study are reported in a doctoral thesis by Anna T. Cianciolo. We gratefully acknowledge the assistance of the Kanfer and Ackerman Laboratory staff for assisting in simulation development, conducting the experiments, and data coding.

Correspondence concerning this article should be addressed to Phillip L. Ackerman, School of Psychology, Mail Code 0170, Georgia Institute of Technology, 274 5th Street, Atlanta, Georgia 30332-0170. E-mail: phillip.ackerman@psych.gatech.edu

to have minimal content-specific demands show this general pattern. For example, the Raven Progressive Matrices test is widely believed (by some) to be a test of pure reasoning ability (Spearman, 1938). However, the stimuli are made up of polygons, lines, dots, and so on, and the test correlates much higher with other tests with spatial or figural content than it does with reasoning tests that depend on the processing of verbal or numerical test items (e.g., see Burke, 1958, for a review). One implication of this suggestion is that surface-level features of tasks may be the most important determining characteristics of ability–task performance relations, and that there may be a dissociation between mean-level indicators of performance and the rank-ordering of individuals in task performance. Such a suggestion is entirely consistent with Kyllonen's (1985) conclusions with basic information-processing tasks but has not been directly assessed with complex tasks.

Whether a task has consistent or inconsistent (variable) mappings of stimuli to responses has been shown to have substantial effects on mean task performance over task practice (e.g., see Shiffrin & Schneider, 1977). That is, when tasks have high overall consistency, large performance improvements result with increased levels of task practice, whereas tasks with low overall consistency tend to show attenuated learning curves. From an individual-differences perspective, Ackerman (1988) suggested that initial performance on both consistent and inconsistent versions of the same task would have equivalent demands on general intellectual abilities. As consistent tasks are learned, a decreasing correlation is found between broad intellectual abilities and task performance, whereas increasing practice may result in no changes to ability–performance correlations for tasks with low consistency. Although many highly consistent tasks have been found to reveal these declining patterns of ability–performance correlations, others have not shown this general pattern (e.g., see Woltz, 1988). Indeed, there is inconsistent evidence concerning whether task content is a more important determinant of ability–performance correlations than is task consistency (e.g., see Ackerman, 1986; Ackerman & Woltz, 1994). For tasks that involve a mixture of consistent and inconsistent mappings of stimuli to responses, it is not clear whether changes in task consistency, per se, increase or decrease ability–performance correlations. We speculate that for such tasks, changes in task consistency, per se, do not result in changes to the correlations between broad content abilities and task performance (except for a situation where consistency reaches an extremely high level, which would allow for the development of automaticity over task practice and thus a reduction in demands on general and broad content abilities). The current investigation attempts to assess whether these findings can be generalized to tasks that involve a substantial initial investment of time and effort toward skill acquisition.

The main goal of the current study is to start with a complex skill-acquisition task, assess individual differences in both relevant broad content abilities and perceptual speed (PS) abilities, and then perform a series of task condition manipulations that are designed to change mean performance and some of the underlying task characteristics. The change in task characteristics is to be evaluated by examination of mean differences between conditions, and the changes in rank order of individual differences is to be evaluated by examination of ability–performance correlations within the initial training phase and during the test conditions. In this way, the robustness of ability predictors of task performance can be evaluated in the context of specific changes to the underlying task characteristics.

To provide a meaningful evaluation of the task characteristic–ability predictions of performance contrast, it is necessary to have a reasonably well-known task platform, that is, sufficiently complex that the task cannot be mastered to a level of asymptotic performance in a few minutes of task practice. It is also necessary to be able to change the task characteristics in a meaningful way that is both germane to the overall task and tractable from a task-analysis perspective. Although one could create two analogous tasks that differ only in content (such as spatial vs. verbal material; see Ackerman, 1986), such a manipulation essentially serves to "snare" the effect of changes to content–ability correlations. Alternatively, one could simply increase the speededness of an extant task, but such a manipulation has been clearly demonstrated to generate a set of speed factors (Carroll, 1993) rather than addressing the underlying capability of substantive ability tests to predict individual differences in performance.

Overview

The current study uses an air traffic control simulation task called TRACON (Terminal Radar Approach Control) that is sufficiently complex that asymptotic performance typically requires about 15 hr of time-on-task (e.g., see Ackerman, 1992; Ackerman & Kanfer, 1993; Ackerman, Kanfer, & Goff, 1995). The TRACON software platform, in several different forms, has been used for training of air traffic controllers for the U.S. Federal Aviation Administration, Department of Defense, several college and university airway sciences programs, and in several foreign countries. In addition, a scaled-down version of the software was a popular simulation–game program during the early-to-mid 1990s. Previous investigations of ability determinants of performance on the task have shown that spatial and numerical abilities, along with complex PS abilities, provide substantial predictive validity for task performance over task practice (e.g., see Ackerman, 1992; Ackerman & Kanfer, 1993). Moreover, one of the most important characteristics of this task, in contrast to many simpler laboratory-based skill-learning tasks, is that TRACON shows stable (or increasing) correlations with general intellectual ability across task practice (for an extensive discussion of this issue, see Ackerman, 1988; Barrett, Alexander, & Doverspike, 1992; Hulin, Henry, & Noon, 1990). That is, the cognitive ability demands of the task do not decline with increasing task practice, a factor that avoids a confound of changing validity coefficients with practice and later task characteristic manipulations.

The TRACON task also has an advantage of differentiated task components in the form of different kinds of flights that are handled in the course of a simulation trial. All planes appear to be similar on the simulated radar screen. There are common elements to handling each kind of flight (e.g., the plane must be accepted into the participant's airspace or "sector," must be guided through the sector without violating the airspace of other planes, and must be properly handed off to another controller before exiting the participant's sector). There are, however, three different kinds of flights to be handled: overflights, arrivals, and departures. An example of the typical commands used for each of these flight types is provided in Table 1.

Table 1
Example of Commands for Arrival, Overflight, and Departure Flight Types

Time (MM:SS)	Command
	Arrival
03:28	App/Dep: UA907, Radar contact.
03:33	App/Dep: UA907, Cleared direct to COCOA.
03:33	UA907: Going direct to COCOA.
03:37	App/Dep: UA907, Change speed to 250.
03:37	UA907: Speeding up to 250.
03:43	App/Dep: UA907, Change altitude to 3,000.
03:43	UA907: Descending to 3,000.
07:49	App/Dep: UA907, Cleared direct to BOJAK.
07:49	UA907: Going direct to BOJAK.
10:59	App/Dep: UA907, Turn left heading 145.
11:00	UA907: Roger. Left to 145.
11:03	App/Dep: UA907, Turn left heading 120.
11:03	UA907: Roger. Left to 120.
11:07	App/Dep: UA907, Cleared direct to MDW.
11:07	UA907: Going direct to MDW.
11:29	App/Dep: UA907, Cleared for ILS approach, . . . Contact tower at FAF.
11:29	UA907: Thanks. Good day.
	Overflight
12:25	App/Dep: N54RF, Radar contact.
22:23	App/Dep: N54RF, Change altitude to 6,000.
22:24	N54RF: Climbing to 6,000.
24:27	App/Dep: N54RF, Change altitude to 4,000.
24:27	N54RF: Descending to 4,000.
25:46	App/Dep: N54RF, Contact center. Good day.
25:46	N54RF: Switching to center frequency.
	Departure
01:54	App/Dep: N83AE, Released.
04:49	App/Dep: N83AE, Change altitude to 7,500.
04:49	N83AE: Climbing to 7,500.
05:32	App/Dep: N83AE, Change altitude to 9,000.
05:33	N83AE: Climbing to 9,000.
18:55	App/Dep: N83AE, Contact center. Good day.
18:55	N83AE: Switching to center frequency.

Note. MM:SS = minutes:seconds; App/Dep = command from air traffic controller; UA907 = United Flight 907; COCOA and BOJAK = radar fixes; MDW = Midway Airport; ILS = instrument landing system; FAF = final approach fix; N54RF and N83AE = general aviation aircraft (small private propeller-driven planes).

Overflights

For an overflight, the participant has the least demands on the information-processing system. When first encountered at a radar fix at one edge of the participant's sector, the plane is already at its cruising altitude and on a direct heading for the exiting radar fix at a different edge of the participant's sector. Under optimal circumstances, the participant only needs to accept the plane into his or her sector, monitor for conflicts, and then hand off the plane within 5 miles of the exiting radar fix. Under less-than-optimal circumstances, the participant must change the direction, altitude, or speed of the plane to avoid other planes and then make sure that the plane is returned to its intended flight path before it is handed off to the appropriate controller.

Arrivals

Arrivals represent the most demanding flights. Arrivals are initially accepted into the sector at a cruising altitude and must be maneuvered to a lower altitude and heading that bring the plane into the tight cone that represents the final approach for the intended airport. This requires that the participant perform at least one or more vectoring commands, along with directing changes in altitude and speed. Arrivals have the greatest likelihood of conflicts with other planes, thus increasing the monitoring activities as the plane changes direction and descends in altitude. Moreover, to line up within a ±5° heading tolerance at the airport, either greater planning or multiple adjustments through heading commands are required, or the plane executes a "missed approach" and must be revectored back to the airport. (As an example, previous data on the TRACON task [Ackerman et al., 1995] indicate that participants make about six times as many maneuver commands, including changes in heading, altitude, and speed, to arrivals than to overflights: $M = 10.65$ for arrivals vs. $M = 1.64$ for overflights.)

Departures

Departures represent a level of complexity somewhere between arrivals and overflights. They are more involved than overflights because once accepted, they take off from an airport and, unless otherwise directed by the participant, turn and climb to reach the cruising altitude and direction to intercept the exit radar fix. As such, the participant must anticipate the changing progress of the flight as it potentially climbs through the airspace of other flights. However, with proper planning, these flights are much less complex than arrivals, because it is not necessary to change the altitude or heading of the flight, as long as potential conflicts are avoided. Consistent with the task analysis of these components, previous experiments (e.g., Ackerman et al., 1995) have indicated that although performance on all types of flights was substantially correlated with spatial, math, and complex PS abilities, performance on the arrival flight components of TRACON had higher correlations with spatial and math abilities than did the overflights. Departure flights tended to have ability correlations somewhere in between those of the arrivals and overflights.

The differences in ability–performance correlations among the three flight types are largely reflected in mean levels of performance. That is, even though previous studies have not emphasized handling one type of flight over another, previous mean performance levels have been higher for handling overflights than for arrivals, with mean number of departures falling in between the other two flight types (e.g., see Ackerman et al., 1995). Thus, at least, based on within-task assessments, there is a concordance between mean levels of performance and magnitude of correlations with abilities, with the easiest task components having lower correlations with content (spatial and math) abilities and PS-Complex ability. Such a pattern of results is consistent with Kyllonen’s (1985) conclusions, although TRACON is more complex than the tasks studied by Kyllonen by at least an order of magnitude.

Task Manipulations and Predictions

The basic experimental design involved first allowing the participants to achieve an adequate level of task performance (accomplished through a 1-hr instructional video and 8.5 hr of time-on-task). Then, a 2 × 2 test condition design was implemented, within-participants, in a counterbalanced order. The first dimension to be manipulated was a shift from a full mixture of flight types (arrivals, overflights, and departures) to either a dominant presentation of arrivals or overflights (with a constant number of departures). That is, with 28 flights in a simulation trial, the arrivals conditions had 20 arrivals and 8 departures, whereas the overflights conditions had 20 overflights and 8 departures. This manipulation can be considered an overall *complexity–content* manipulation, in that arrival flights are more complex to successfully complete and also that arrival flight performance is expected to have concomitant higher demands on spatial and general intellectual abilities, as compared with overflight performance (e.g., see Ackerman, 1992; Ackerman et al., 1995).

The second dimension to be manipulated was a change in the compliance of planes to participant-issued commands. That is, in the training trials, the planes always complied with participant-issued commands (as long as they were legitimate commands)—a 100% compliance. In the compliance manipulation, participants were either confronted with 100% compliance or a reduced compliance condition (65% compliance), where the probability that the plane would correctly execute the participant-issued command was reduced to an average 65% of all issued commands. This manipulation is analogous to a reduction of participant input–system output consistency (see e.g., discussions of task consistency manipulations on ability–performance relations by Ackerman, 1986, 1987, 1988; Ackerman & Woltz, 1994). The arrivals versus overflights manipulation was designed to either increase or decrease, respectively, the overall task complexity and the concomitant cognitive–intellectual demands of the task. The 65% compliance condition was designed to increase the overall task difficulty, especially by increasing the amount of task monitoring required during the simulations. We expected main effects for each of the two manipulations and a strong interaction effect (given that arrivals involve many more commands, on average, than overflights, such that increasing numbers of commands and subsequent monitoring are required by the arrivals–65% compliance condition, in comparison with the overflights–65% condition).

Although the predictions of mean performance differences among the conditions are straightforward and noncontroversial, the predictions of changes to ability–performance relations represent the crux of the theoretical and substantive questions. Those investigators who believe that there should be an isomorphic correspondence between task information-processing demands and ability–performance relations would predict that the correlations between abilities and performance should track the mean levels of performance. That is, more difficult or complex task conditions lead to lower levels of performance and thus should result in higher correlations between performance and ability, as long as there is no confounding restriction of range associated with floor effects. Conversely, easier task conditions lead to higher levels of performance and thus should result in lower levels of correlations between performance and ability, as long as there is no restriction of range associated with ceiling effects. If Kyllonen’s (1985) findings of basic information-processing tasks generalize, it may be that content, task complexity, and task difficulty affect the ability–performance correlations. In contrast, if Ackerman’s (2000) hypothesis is correct, then the overall content of the task is the driving determinant of individual differences in task performance. As such, it might be predicted that only changes to the content of the task would result in changes to ability–performance relations. That is, contrasting predominantly arrival trials with predominantly overflight trials, which is a task content manipulation, results in a change to ability–performance correlations, but contrasting the 100% and 65% compliance conditions shows no change in ability–performance correlations (because the content of the tasks is unchanged, even though the difficulty of the tasks may change). We expected that the latter set of predictions would be borne out in the analysis of the task performance. That is, (a) no changes in ability–performance correlations for the compliance conditions (which is essentially an expectation of affirming the null hypothesis—and as such, would require an assessment of the statistical power of the assessment, should the null hypothesis not be rejected), and (b) predominantly arrival flight simulations would show higher correlations with general and PS abilities than predominantly overflight flight simulations.

Method

Participants

Eighty-one participants between the ages of 18 and 30 years participated in this study (mean age = 21.1 years, *SD* = 2.53). There were 36 women and 45 men. All participants were native English speakers, current college–university students or graduates, and had normal or corrected-to-normal vision, hearing, and motor coordination.

Apparatus

Paper-and-pencil testing. Ability tests were administered to groups of up to 16 participants at a time in a classroom-like setting. Test instructions and start–stop times were presented over a public address system on prerecorded minidiscs.

Task apparatus. Training and test trials for the air traffic controller task were administered at individual carrels, with participants separated by partitions. The task was conducted for up to 16 participants at a time. The task was administered on Dell and IBM Pentium computers with Trinitron 17-in. (43.2-cm) monitors. Visual information was displayed in color VGA (640 horizontal pixels × 480 vertical pixels) resolution graphics. Audio information from the TRACON task was presented binaurally through headphones, using a SoundBlaster (Creative Technology, Singapore, China) interface. Participants interacted with the task with standard IBM PC 101 keyboards and a MicroTech (SCM Microsystems, Guilford, CT) three-button trackball.

Ability Measures

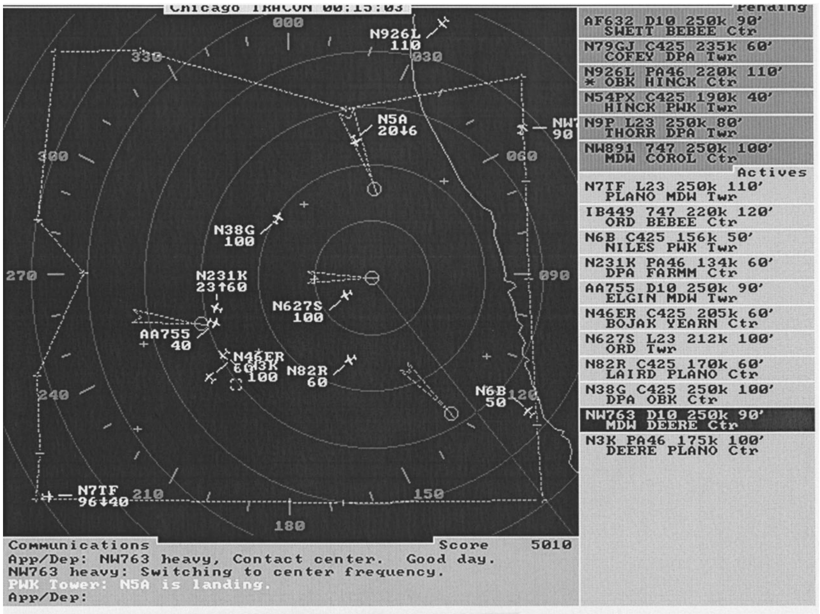
Eighteen tests were administered to obtain estimates of two broad content abilities (Spatial and Numerical) and four PS factors (PS-Complex, PS-Memory, PS-Pattern Recognition, and PS-Scanning). With the exception of the PS-Complex factor, at least three tests were administered for each hypothesized underlying ability factor. The tests associated with each factor are listed below (for a full discussion of the PS tests, see Ackerman & Cianciolo, 2000, and the additional references listed for each test below):

> *Spatial ability:* (a) Spatial Analogy, (b) Spatial Orientation, (c) Paper Folding, and (d) Verbal Test of Spatial Ability (all tests are fully described in Ackerman & Kanfer, 1993).
>
> *Numerical ability:* (a) Math Knowledge, (b) Number Series, and (c) Problem Solving (see Ackerman & Kanfer, 1993).
>
> *PS-Memory:* (a) Coding, (b) Digit/Symbol (e.g., see Wechsler, 1958), and (c) Naming Symbols.
>
> *PS-Pattern Recognition:* (a) Canceling Symbols, (b) Finding ∈ and ¥, and (c) Finding a and t.
>
> *PS-Scanning:* (a) Name Comparison (e.g., see Andrew & Paterson, 1934), (b) Number Comparison (e.g., Andrew & Paterson, 1934), and (c) Clerical Abilities–2 (Bennett & Gelink, 1951).
>
> *PS-Complex:* (a) Directional Headings (Cobb & Mathews, 1972; see also Ackerman & Kanfer, 1993), and (b) Dial Reading (Guilford & Lacey, 1947; see also Ackerman & Kanfer, 1993).

TRACON

The software used in this experiment is a modification of the early professional version (V1.52) of TRACON, developed by Wesson International, that allows for collection of a variety of data. The following description of the TRACON platform and task trial design is reprinted from Ackerman and Kanfer (1993).

> A full description of the TRACON simulations is provided in Ackerman (1992). The following discussion abstracts that material, with deviations in procedures specifically noted. The task requires that trainees learn a set of rules for air traffic control, including (a) reading flight strips; (b) declarative knowledge about radar beacons, airport locations, airport tower handoff procedures, and enroute center handoff procedures; (c) airplane separation rules and procedures; (d) monitoring strategies; and (e) strategies for sequencing airplanes for maximum efficient and safe sector traversal. In addition, trainees are required to acquire human-computer interface skills, including issuing trackball-based commands, menu retrieval, keyboard operations, and integration between visual and auditory information channels. Note, however, that the simulator task represents a substantial reduction of rules and operational demands in comparison with the real-world job of [air traffic controllers].
>
> *Display.* TRACON presents the trainee with a simulated color radar screen, depicting a region of airspace, very high frequency omnidirectional range stations (VOR), airports, sector boundaries, and range rings. Planes are identified by an icon on the radar scope, with a data tag that indicates plane identification and altitude information. In addition, two sets of "flight strips" are presented at the right side of the display, a "pending" and an "active" set. Each flight strip contains information about a particular flight, including identification information, plane type, requested speed and altitude, and sector entry-and-exit destination information (see Figure 1). Finally, at the bottom of the screen is a "communications box," which shows commands issued to planes (and responses by the pilots), along with the controller's "score" for the current simulation. When planes were about to enter the trainee's sector (at a boundary or on the runway of an airport), this information was announced over the headset. No flight was allowed to cross the sector boundary or take off from an airport without explicit authorization by the trainee.
>
> *Task controls and knowledge of results.* Trainees interacted with the TRACON simulation in several ways. A trackball was used for the majority of input activities, although the keyboard was also used alone, or in conjunction with the trackball (the trackball represents a change in input device from the previous study, where a mouse was used). For each plane command, a menu of command choices was displayed on the screen.
>
> Knowledge of results was provided visually (by text in the communications box) and aurally with a read-back by the pilot or other controller (using digitized speech). In addition, planes followed (as nearly as possible) the commands issued by the trainee. Turn, altitude change, and speed change commands were processed by the computer and were carried out in accordance with the limitations imposed by each aircraft type.
>
> When errors occurred (e.g., separation conflicts, near misses, crashes, missed approaches, handoff errors), additional information was presented to the trainee. In each of these cases, an alert circle around the plane(s) in question was presented on the screen, and a series of tones were presented over the headset.
>
> *Trial description.* Trials for the task were created and pretested to be roughly equivalent in difficulty. Each trial contained planes that were divided into three basic categories (overflights, departures, and arrivals). Overflights were planes that entered and exited the trainee's airspace at cruising altitudes. Trainees were required to acknowledge these airplanes as they approached a boundary VOR fix, monitor progress through the sector, and handoff to a "center" controller. Departures were planes that originated at one of the four airports, climbed to a cruising altitude, and were handed off to a center controller. Trainees were required to release departures from airports, evaluate and remediate potential conflicts as the planes climbed to a cruising altitude and turned to intercept their intended flight paths, and then handoff planes to the appropriate center controller. Arrivals entered the trainee's airspace from one of the boundary VOR fixes and had to be landed at a designated airport. Trainees were required to direct arrivals onto an appropriate heading and altitude to provide an acceptable handoff to the appropriate "tower" controller, then these



```
Pending
AF632 D10 250k 90'
  SWETT BEBEE Ctr
N79GJ C425 235k 60'
  COFEY DPA Twr
N926L PA46 220k 110'
* OBK HINCK Ctr
N54PX C425 190k 40'
  HINCK PWK Twr
N9P L23 250k 80'
  THORR DPA Twr
NW891 747 250k 100'
  MDW COROL Ctr
Actives
N7TF L23 250k 110'
  PLANO MDW Twr
IB449 747 220k 120'
  ORD BEBEE Ctr
N6B C425 156k 50'
  NILES PWK Twr
N231K PA46 134k 60'
  DPA FARMM Ctr
AA755 D10 250k 90'
  ELGIN MDW Twr
N46ER C425 205k 60'
  BOJAK YEARN Ctr
N627S L23 212k 100'
  ORD Twr
N82R C425 170k 60'
  LAIRD PLANO Ctr
N38G C425 250k 100'
  DPA OBK Ctr
NW763 D10 250k 90'
  MDW DEERE Ctr
N3K PA46 175k 100'
  DEERE PLANO Ctr

Communications                                    Score    5010
App/Dep: NW763 heavy, Contact center.  Good day.
NW763 heavy: Switching to center frequency.
PWK Tower: N5A is landing.
App/Dep:
```

Figure 1. Static copy of [terminal radar approach controller] TRACON screen. There are three major components to the display. The right-hand side of the screen shows pending (not under control) and active (under control) flight strips. Each flight strip lists (a) plane identifier, (b) plane type, (c) requested speed, (d) requested altitude, (e) radar fix of sector entry, and (f) radar fix of sector exit (including tower or center). The lower part of the screen shows a communications box that gives a printout of the current (and last few) commands issued by the trainee and the responses from pilots or other controllers. The main part of the screen shows a radar representation of the Chicago sector. Planes are represented by a plane icon and a data tag, which gives the identifier, the altitude, and an indication of current changes in altitude. The sector is bounded by the irregular dotted polygon describing a perimeter. Radar fixes are shown as small (+) figures on the radar screen. Airports are shown with approach cones and a circle indicating the facility proper. A continuous radar sweep is shown (updating at 12 o'clock every 5 s). Range rings are also displayed indicating 5-mile (8.0467-km) distances. From "Cognitive and Noncognitive Determinants and Consequences of Complex Skill Acquisition," by P. L. Ackerman, R. Kanfer, and M. Goff, 1995, *Journal of Experimental Psychology: Applied, 1,* p. 303. Copyright 1995 by the American Psychological Association. Reprinted with permission of the author.

planes would land. For all flights, the trainee was required to maintain legal separation (at least 1,000 ft [304.8 m] in altitude, or 3 mi [5.56 km] horizontally).

A successful "handle" of a flight was the appropriate accomplishment of the respective flight plan. That is, for a departure or an overflight, the accomplishment was a successful handoff to the appropriate center controller. For a landing, the accomplishment was the successful landing of the airplane.

Performance measurement. As with a previous investigation (Ackerman, 1992), overall performance was computed as the sum of all flights accepted into the sector that had a final disposition within the simulation time (minus any planes that were incorrectly disposed of, e.g., crashes, not-handed-off, vectored off the radar screen). This measure is concordant with measures derived from the examination of the criterion space for FAA ATC simulation research (e.g., see Buckley, DeBaryshe, Hitchner, & Kohn, 1983) and has acceptable reliability.

TRACON training trials. Each [training] trial was comprised of 16 overflights and departures (with roughly equal frequency), and 12 arrivals. The airplanes requested entry to the airspace at irregular intervals that were constrained to require the trainee to be always occupied with at least one active target. The trials were also constrained so that perfect performance (handling all 28 airplanes successfully) was just beyond the skill level achieved by subject matter experts. Each trial was concluded in 30 min (Ackerman & Kanfer, 1993, p. 417).

TRACON Test Trials

There were four types of TRACON test trials in a 2 (arrivals only vs. overflights only) × 2 (100% compliance vs. 65% compliance) design. The 100% pilot compliance procedure was the same as in training (i.e., the pilots always responded to the commands issued unless they were incapable of compliance; e.g., a command to descend below minimum altitudes or to be handed off to a center controller with greater than 5 miles from the edge of the control sector would receive a "cannot comply" response).

The first manipulation was to shift the task parameters so that the participants received a majority of either arrival flights or overflights. These conditions were called arrivals only and overflights only even though each simulation trial also contained a control set of eight departure flights. In the arrivals-only conditions, there were 20 arrival flights but no overflights. In the overflights-only conditions, there were 20 overflights and no arrival flights.

The second manipulation was to contrast 100% compliance by pilots with a condition where pilot compliance was degraded—called the 65% compliance condition. The 65% compliance conditions were designed so that, on average, the pilots respond appropriately only 65% of the time. That is, roughly 35% of the time, pilots either failed to respond to or failed to comply with the participant's commands. For example, the pilot might reply "Repeat that. I was looking at a map." In such circumstances, the participant must reissue the command, taking account of any change in plane location in the interim between the initial command and the follow-up command. In some cases when the pilot failed to reply, another pilot may comply with the command (e.g., a command issued to "American, Flight 135" to "turn right to 270°" might get a complying response from "United, Flight 27" so that the American Flight 135 does not turn but the United Flight 27 does turn right to 270°. Under such conditions, the participant would need to issue the command again to the American flight and also correct the erroneous action by the United flight. Auditory read-back provided to the participant was consistent with any actions taken (e.g., if there was no response from the pilot, there would be no auditory or text response in the command window; and if there was a response from the wrong pilot, that information would be displayed and presented over the headsets). To perform well under these conditions, the participant needed to attend to both the intended command and the actual read-back or response, note any discrepancies, and quickly issue commands to remediate the discrepancies.

Manipulation Check

At the completion of the final test trials, a brief questionnaire was administered. The questionnaire first described the last TRACON training trials and the four different test conditions and then asked for a subjective difficulty rating: "For the ______ trials, I found the TRACON trials to be ______," where the first blank referred to condition type (e.g., "Mostly overflights, pilot compliance 100%") and the second blank referred to a provided scale ranging from 0 (*very easy*) to 100 (*extremely difficult*).

Procedure

This study was completed in seven sessions, totaling approximately 26 hr. Session 1 was devoted entirely to pencil-and-paper ability testing (some of which were part of a larger study and the handout of a take-home questionnaire, neither of which are reported here). Session 2 began with a TRACON instructional video, followed by practice on TRACON simulations. The rest of Session 2 was devoted to 90 min of TRACON training trials (three 30-min trials). TRACON trials were always 30 min in length, and 5-min breaks were given after every two trials. Sessions 3 and 4 were devoted entirely to TRACON training trials (six trials each). Session 5 started with two final TRACON training trials. The remainder of the session was devoted to the completion of four 30-min TRACON test trials. Before starting this session, participants received the following instructions:

> During this session, after your first break, you will notice some changes in the simulations. Some of them will have more arrivals than the ones on which you have practiced. Some will have more overflights. In addition, you will notice that sometimes the pilots' responses to your commands will be different. These are not difficulties with the program but are intentional modifications to the simulations. That is, in these trials you may have to reissue commands and plan ahead for pilot error. When I tell you to begin, press the ⟨*Spacebar*⟩ key on the keyboard and the task will start.

Sessions 6 and 7 were devoted entirely to the completion of the remaining TRACON test trials (six trials each). At the end of Session 7, participants were debriefed and remunerated $250 each—not contingent on performance.

A total of 17 TRACON training trials were thus completed (8.5 hr time-on-task), along with 16 test trials: 8 hr of test trials, 4 trials each for the 2 (arrival vs. overflight) × 2 (100% compliance vs. 65% compliance) conditions. The order of both the TRACON training and test trials was counterbalanced across participants in a Latin-square design. For the test trials, the counterbalanced order presented each participant with alternating arrivals versus overflights trials. In addition, the degree of pilot compliance was alternated every other trial. Finally, the same set of simulations assigned to the 100% pilot compliance condition for half the participants was assigned to the 65% average pilot compliance condition for the other half of the participants.

Results

The results are presented in six sections. The first section focuses on the derivation of the reference ability factors for later computation of ability–performance correlations. The second section reviews the training data from the first seventeen 30-min trials of TRACON. The third section concerns the effects of the arrival versus overflight and 100% versus 65% compliance manipulations on task performance—both for component flight types and for overall performance. The fourth section reviews self-report estimates of task difficulty for the final training and the test conditions. The fifth section describes the correlations between TRACON training trial performance with the two content abilities (spatial and numerical) and the four PS factors (PS-Complex, PS-Memory, PS-Scanning, and PS-Pattern Recognition), along with more molar correlations with a general ability composite and a general PS composite. The sixth and final section addresses the correlations between ability composites and task component and overall performance for the four TRACON test conditions.

Ability tests and factors. Because the reliability and validity of these measures have all been previously reported in the literature (e.g., see Ackerman, Bowen, Beier, & Kanfer, 2001; Ackerman & Cianciolo, 2000; Ackerman & Kanfer, 1993; Ackerman et al., 1995), we only briefly review the initial analysis of the ability measures. Means, standard deviations, and intercorrelations of the 18 ability tests are presented in Table 2. The scores on these tests were largely comparable to those obtained with other college–university student samples, with some indication that this sample had higher overall scores on math and spatial ability tests (a result consistent with the fact that these participants were sampled from a student population that is somewhat more selective than that of the previous studies). In comparison with the sample of University of Minnesota students reported in Ackerman et al. (1995), for the tests that overlapped between the two studies, the current sample had mean performance on math and spatial tests $.55\sigma$ higher than the previous sample.

Because the content of the PS tests (e.g., numerical and spatial stimuli) overlapped in several cases with broad content tests, the content tests and PS tests were factor analyzed separately. This avoids both a collapsing of the subordinate PS factors into a single overall PS factor and mixing the content factors with the highly speeded PS tests. (For an extensive discussion of this issue, see Ackerman & Cianciolo, 2000.) Both factor analyses were derived with squared multiple correlations as initial communality estimates and a Humphreys–Montanelli parallel analysis (see Humphreys & Montanelli, 1975; also see a more recent treatment by Keeling, 2000) to determine the number of underlying factors. Factors were then rotated to simple structure via Tucker's Direct Personal Probability Factor Rotation (Tucker & Finkbeiner, 1981), using Carroll's factor analysis software suite (Carroll, 1989). Respective factor patterns and factor intercorrelations for the content tests and PS tests are presented in Table 3. In each case, a priori expecta-

Table 2
Means, Standard Deviations, and Intercorrelations for Ability Tests

Variable	*M*	*SD*	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. Spatial Analogy	17.90	6.14	—																	
2. Spatial Orientation	9.50	4.48	.566	—																
3. Paper Folding	15.19	5.41	.650	.550	—															
4. Verbal Test of Spatial Ability	13.81	5.28	.463	.642	.545	—														
5. Math Knowledge	22.71	7.45	.412	.318	.406	.408	—													
6. Number Series	10.75	3.09	.493	.223	.483	.345	.480	—												
7. Problem Solving	5.61	2.81	.491	.385	.506	.509	.669	.478	—											
8. Coding	106.44	24.43	.355	.187	.175	.302	.243	.358	.257	—										
9. Digit/Symbol	233.83	48.70	.247	.285	.148	.357	.095	.118	.011	.603	—									
10. Naming Symbols	306.59	61.35	.339	.228	.193	.278	.267	.328	.243	.586	.682	—								
11. Canceling Symbols	118.54	30.28	.181	.202	.318	.246	.028	.258	.062	.362	.242	.295	—							
12. Finding ∈ and ¥	126.47	27.24	.197	.019	.163	.098	−.089	.195	−.107	.378	.354	.377	.604	—						
13. Finding a and t	156.73	32.07	−.034	−.020	−.008	.068	−.136	.078	−.146	.437	.373	.350	.611	.720	—					
14. Name Comparison	79.40	18.98	.354	.111	.153	.288	.266	.386	.235	.576	.372	.423	.327	.437	.343	—				
15. Number Comparison	82.38	18.00	.207	−.001	.095	.114	.180	.371	.066	.553	.375	.438	.474	.547	.558	.727	—			
16. Clerical Abilities–2	36.85	8.65	.455	.245	.207	.429	.423	.489	.418	.522	.420	.496	.217	.260	.340	.585	.519	—		
17. Directional Headings	91.44	29.41	.584	.466	.491	.600	.439	.481	.465	.493	.512	.479	.328	.295	.265	.498	.349	.638	—	
18. Dial Reading	28.66	8.13	.644	.383	.548	.501	.541	.540	.610	.335	.299	.361	.260	.288	.086	.435	.314	.626	.706	—

Note. For *r*s $> .218$, $p < .05$; for *r*s $> .295$, $p < .01$.

tions were largely confirmed (two factors underlying the content tests—Spatial and Numerical; and four factors underlying the PS tests—PS-Memory, PS-Pattern Recognition, PS-Scanning, and PS-Complex). Only one variable had a significant loading on another factor than initially hypothesized (i.e., the Clerical Abilities–2 test had a significant loading on the PS-Complex factor in addition to the significant loading on PS-Scanning). For both the content factor solution and the PS factor solution, the factors showed positive manifold, indicating that general factors in each case were also implied. We return to these general factors in the treatment of criterion-related validity. On the basis of confirming information from the factor analyses, we created composites for each of the ability factors, using summed unit-weighted *z* scores (e.g., see Thorndike, 1986).

TRACON performance: Training trials. For the 17 training simulation trials, mean performance levels (planes handled) overall and for the three component flight types (arrivals, overflights, and departures) are shown in Figure 2. Repeated measures analyses of variance (ANOVAs) were performed on each of these measures (though the ANOVA on overall performance is redundant with the component flight types). Consistent with previous studies, and as is clear from the figure, there was a strong effect of task practice on performance for each of the three flight types, $F(16, 1264) = 79.95$, 21.58, and 18.48 for arrivals, overflights, and departures, respectively; and for overall performance, $F(16, 1264) = 83.21$, with effects significant at $p < .01$. In previous investigations (e.g., Ackerman et al., 1995), performance improvements were seen through 30 simulation trials. The patterns of improvement in task performance were consistent with the notion that asymptotic performance had not yet been achieved by the 17th task trial for any of the component flight types. Polynomial regressions for trial number were consistent with this interpretation. In each case, a large and significant correlation was found for a linear improvement across trials ($r_{\text{linear}} = .96$, .96, and .97, $p < .01$) for arrivals, overflights, and departures. Small and nonsignificant coefficients were found for the quadratic trends ($r_{\text{quadradtic}} = .22$, .19, and .18). Coefficients for the cubic trend were even smaller; none exceeded $r = .10$. Thus, although performance improved substantially over the course of the training trials (for overall performance, $M = 8.14$, $SD = 4.79$ for Trial 1; and $M = 18.91$, $SD = 6.88$ for Trial 17), the performance curves show continued learning was occurring, even after 8.5 hr of time-on-task (not to mention the additional hour of video instruction before the initial task trial). It is, though, useful to note that by the 17th trial, 7 (9%) of the participants had reached perfect performance. In comparison with previous study results (e.g., Ackerman et al., 1995), where 15 hr of time-on-task practice was provided to 93 participants, the mean for final task performance was 19.49 ($SD = 6.57$). A *t* test for the difference between means provided the following result, $t(172) = -0.57$, *ns;* $d = -0.09$, indicating that there was no significant difference between the performance of the two groups at the end of practice. Such a result is consistent with the higher spatial and math abilities of the current sample of participants and suggests that this particular group performed as well at 8.5 hr of practice as a less able group did at 15 hr of practice.

Although it is possible that participants traded off performance among the task components (e.g., by focusing only on overflights, to the exclusion of arrivals), which would result in a negative

Table 3
Ability Test Factor Patterns and Intercorrelations: DAPPFR Rotations

	Factor	
Test	Spatial	Numerical
Spatial Analogy	**.430**	.190
Spatial Orientation	**.803**	−.231
Paper Folding	**.454**	.183
Verbal Test of Spatial Ability	**.515**	.070
Math Knowledge	.000	**.582**
Number Series	.049	**.472**
Problem Solving	.098	**.553**
Correlations		
Spatial	—	—
Numerical	.639	—

	Factor			
	PS-Memory	PS-Pattern Recognition	PS-Scanning	PS-Complex
Coding	**.410**	.042	.268	−.005
Digit/Symbol	**.749**	−.018	−.049	−.016
Naming Symbols	**.562**	.014	.066	.052
Canceling Symbols	−.049	**.620**	−.031	.108
Finding ∈ and ¥	−.009	**.678**	.007	.047
Finding a and t	.070	**.674**	.025	−.114
Name Comparison	.016	.014	**.610**	.089
Number Comparison	−.025	.249	**.574**	−.081
Clerical Abilities-2	.190	−.063	**.303**	**.364**
Directional Headings	.280	.022	.019	**.550**
Dial Reading	.020	.023	.010	**.806**
Correlations				
PS-Memory				
PS-Pattern Recognition	.487	—	—	—
PS-Scanning	.514	.504	—	—
PS-Complex	.354	.219	.431	—

Note. Factor loadings > .30 are shown in boldface. DAPPFR = Direct Artificial Personal Probability Factor Rotation (Tucker & Finkbeiner, 1981).

correlation among task components, the data do not support this particular speculation. For the first task trial, the average correlation among the three task components was .473. At the end of practice (Trial 17), the average correlations among task components had increased to .576. One cannot directly disconfirm the notion that some individual participants may have traded off performance. However, the overall positive association among these task components suggests that individuals who performed well on one component also performed well on the other task components.

TRACON performance: Test trials. Administration of the test trials followed the counterbalanced order described in the Method section. Performance on the four trials for each condition were averaged. Given that the mixture of flight types differed between the arrival and overflight conditions, overall mean performance was not directly contrasted. Instead, separate analyses were performed for the compliance conditions for each type of trial. Table 4 shows means and standard deviations for each of these conditions, along with a test for the effects of the compliance conditions within arrival and overflight trials.

In both types of trials, as expected, reduction of pilot compliance from 100% to 65% resulted in a significant reduction in performance. For overflights, the effect of the reduction in compliance was $d = 0.64$, that is, a difference between means of .64 standard deviation units (see Cohen, 1988, for a discussion of effect sizes). For the arrival conditions, the effect was of a similar form but over twice as large, $d = 1.62$, indicating that the effect of reduced compliance was much larger for the arrival trials than for the overflight trials.

Because departure flights represented a within-task control (each of the overflight and arrival trials had the same number of departure flights), it is possible to examine these data in a fully crossed repeated-measures ANOVA (though the separate analyses within conditions are also presented for comparison purposes). The ANOVA results are shown at the bottom of Table 4. In that analysis, all of the main and interaction effects were significant. Examination of the means and t tests, however, makes interpretation much simpler. That is, significantly fewer departures were handled in both arrival conditions, and the reduction of compliance from 100% to 65% was only significant or substantial in the

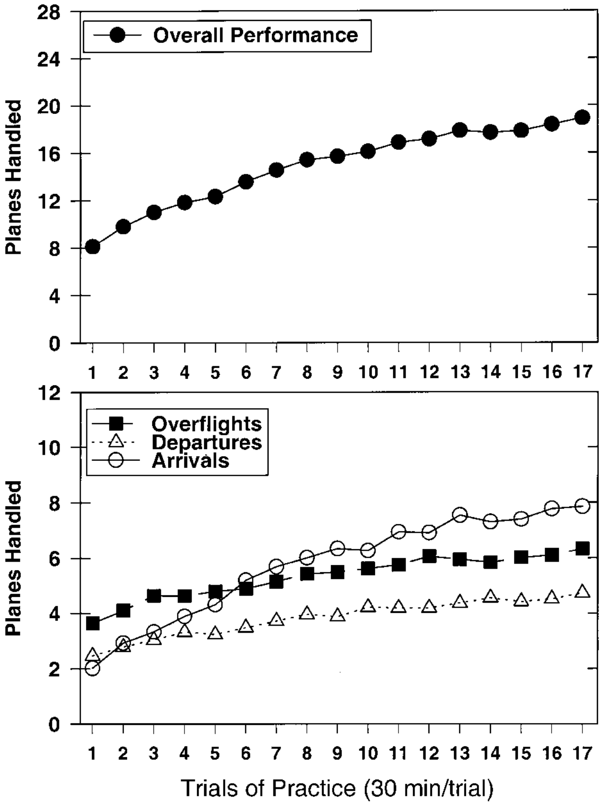


Figure 2. Mean performance (planes handled) during training trials (top). Overall performance (arrivals + overflights + departures). Planes handled by flight type (bottom). Maximum possible performance was 12 arrivals and, on average, 8 overflights and 8 departures per trial, or a total of 28 planes overall.

arrivals condition. Conversely, reduction in compliance had virtually no effect in the overflight condition ($d = -0.03$).

The overall picture that emerges from the test condition performance data consist of the following two major points:

1. Task performance was largely degraded by the introduction of the 65% compliance condition but was especially degraded in the arrivals trials, as compared with the overflights trials.

2. Concomitant effects on departures were generally consistent with the patterns for arrivals and overflights, except that the reduced compliance had virtually no effect on departures handled in the overflight trials.

Manipulation check. The posttask questionnaire provided descriptions of each of the test conditions and asked participants to rate difficulty. Analysis of the posttask questionnaire data (see Table 5) indicates that subjective difficulty was significantly higher for the arrivals conditions and was also significantly higher for the 65% compliance conditions than for the 100% compliance conditions. Examination of the significant interaction between arrivals–overflights and 100%–65% compliance conditions indicated that, similar to the findings of departure flight performance, the most difficult condition was the arrivals–65% compliance condition. The mean difficulty rating for this condition corresponded to a judgment that the task condition was “difficult.”

In comparison with the ratings of the final TRACON training trials, subjective difficulty of the test condition trials was significantly lower for the overflight–100% condition and the arrivals–100% condition, $t(80) = 5.85$ and 3.76, respectively, $p < .01$. Difficulty was significantly higher for the arrivals–65% condition, $t(80) = -5.30$, $p < .01$, and not significantly different from the final training trials for the overflight–65% condition, $t(80) = -0.91$, *ns.*

In addition to the self-reported difficulty, an analysis of the average number of commands issued to flights in each of the four conditions provides support for the increased number of activities required by the lower levels of compliance. In the arrivals condition, the mean number of heading, speed, and altitude maneuver commands per flight increased from $M = 8.64$, $SD = 3.93$ in the 100% compliance condition to $M = 13.22$, $SD = 6.52$ in the 65% compliance condition, a significant difference, $t(79) = -11.87$, $p < .01$, $d = 0.85$. Increases of an even larger comparative magnitude were found for the overflight conditions ($M = 2.17$, $SD = 0.67$ in the 100% compliance condition; and $M = 3.03$, $SD = 1.30$ in the 65% compliance condition), $t(79) = -8.49$, $p < .01$, $d = 0.95$.

Ability determinants of TRACON training performance. To evaluate the associations between abilities and training task performance, the two content ability composites (Spatial and Numerical) and the four PS ability composites (PS-Complex, PS-Memory, PS-Scanning, and PS-Pattern Recognition) were correlated with overall task performance. To provide additional stability through aggregation of the criterion data, individual trial performance was averaged into groups of three 30-min trials (and for the last group of trials, a group of two trials). The correlations between the ability composites and task performance during training are shown in the top panel of Figure 3. Consistent with a priori expectations, and largely similar to previous results with this task, correlations with Spatial and Numerical abilities showed initially increasing correlations (with $r = .37$ and .46, respectively, for the first group of trials) with a relatively rapid rise to an asymptote around .60 and .65 for Numerical and Spatial abilities, respectively. As can be seen in the figure, virtually the same pattern of correlations was found for these two abilities. (An issue that often arises in the evaluation of TRACON performance relates to gender differences.) Significant and substantial gender differences were found in task performance for this study (on the order of 1σ *SD* difference in means), as with earlier studies (e.g., see Ackerman et al., 1995), consistent with the demands of the task on spatial and numerical abilities and the general finding of gender differences on these two broad classes of abilities. However, the study of gender differences is not a focus of the current study, and as such, is not further considered here.

The pattern of correlations over practice for the PS factors, shown in the lower panel of Figure 3, was more differentiated than it was for the content abilities. The PS-Complex composite showed large ($r > .50$) correlations throughout practice (with initial increases and later stable correlations), PS-Scanning showed more modest correlations ($r = .30$) throughout practice, whereas PS-Memory showed declining correlations, and PS-Pattern Recognition showed nonsignificant correlations throughout practice.

To focus on the major results, we focused remaining analyses on a more molar level of analysis. That is, a general ability composite was formed from the Spatial and Numerical abilities, and a general PS composite was formed from the four PS abilities. (This aggregation is consistent with the positive manifold found in the two

Table 4
Summary of Test Conditions Analysis

Variable	100% compliance	65% compliance	*t*	*d*
"Overflight" conditions				
Overflights handled				
M	17.75	16.78	5.70**	0.64
SD	2.62	2.58		
"Arrivals" conditions				
Arrivals handled				
M	14.39	11.55	14.51**	1.62
SD	5.10	4.62		
Departures handled				
"Overflight" conditions				
M	5.85	5.88	−0.27	−0.03
SD	2.17	2.07		
"Arrivals" conditions				
M	4.80	3.68	9.24**	1.03
SD	2.39	2.20		

ANOVA for departures handled	*df*	*MS*	*MSE*	*F*	*d*
Arrivals vs. overflights	1, 79	212.88	1.38	154.56**	2.80
100% vs. 65% compliance	1, 79	24.20	0.39	61.95**	1.77
Interaction	1, 79	26.45	0.54	48.75**	1.57

Note. One participant had missing data for one condition, thus a 1 *df* reduction. For Cohen's *d* computation, see Rosenthal, Rosnow, and Rubin (2000, Equation 2.8). ANOVA = analysis of variance.
** $p < .01$.

respective factor analyses discussed earlier.) In this way, we can address the separate components of the TRACON task (i.e., arrivals, overflights, and departures) with just two major predictors, General ability and PS ability. Overall associations between flight types and abilities are shown in Table 6. By aggregating across training trials, it is clear that arrival flights were more highly associated with General ability ($r = .751$) than were either overflights ($r = .597$) or departures ($r = .581$). A t test for the difference between correlations yielded a significant contrast between general ability with arrivals and overflights, $t(78) = 3.04$, $p < .01$, and between arrivals and departures, $t(78) = 3.04$, $p < .01$; and no significant difference between overflights and departures, $t(78) = 0.28$, *ns*. In contrast, no significant differences were found between the flight types and a general PS ability. Thus,

Table 5
Test Conditions for Self-Reported Difficulty

Variable	100% compliance	65% compliance	*t*	*d*
"Overflight" Conditions				
M	27.12	46.56	−9.79**	−1.09
SD	21.62	22.80		
"Arrivals" Conditions				
M	33.14	58.84	−10.24**	−1.14
SD	21.35	20.65		

ANOVA for self-reported difficulty	*df*	*MS*	*MSE*	*F*	***f***
Arrivals vs. overflights	1, 80	6778.78	769.33	8.81**	0.66
100% vs. 65% compliance	1, 80	41254.12	349.56	118.02**	2.43
Interaction	1, 80	796.49	65.24	12.21**	0.78
Last training trials					
M		43.88			
SD		20.88			

Note. For Cohen's *d* computation, see Rosenthal, Rosnow, and Rubin (2000, Equation 2.8). Response scale 0–100 (0 = very easy, 20 = easy, 40 = somewhat difficult, 60 = difficult, 80 = very difficult, 100 = extremely difficult). ANOVA = analysis of variance; ***f*** = Cohen's *f* effect size.
** $p < .01$.

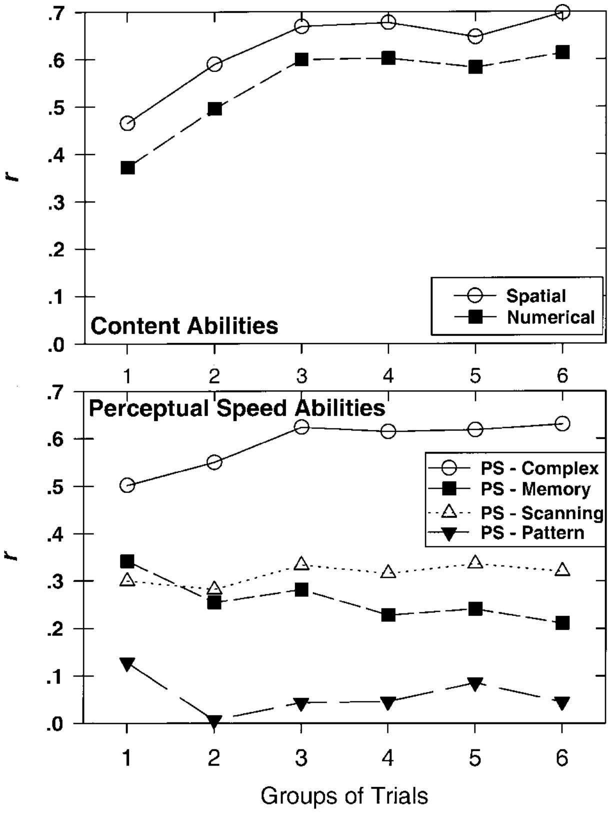


Figure 3. Correlations between ability composites and overall Terminal Radar Approach Control performance during training trials. (Except for Group 6, each point represents an average of three trials. Group 6 is an average of two trials.) For correlations > .218, $p < .05$; for correlations > .295, $p < .01$. PS = perceptual speed.

Table 6

TRACON Component Reliabilities and Correlations Between TRACON Training Performance and General Ability Composites

Variable	1	2	3	4	5	6
Aggregate TRACON training performance						
1. Arrivals	.974[a]	—				
2. Overflights	.779	.956[a]	—			
3. Departures	.732	.793	.964[a]	—		
4. Overall	.938[b]	.916[b]	.897[b]	.982[a]		
General ability composite						
5. General	.751	.597	.581	.717	—	
6. PS	.410	.373	.379	.425	.522	—

Note. For correlations > .218, $p < .05$; for correlations > .295, $p < .01$. TRACON = Terminal Radar Approach Control; PS = perceptual speed.
[a] Within-component reliabilities are Cronbach's alpha internal consistency reliabilities, calculated across the 17 practice trials. [b] Correlations shown are part–whole correlations.

consistent with expectations and the extant literature, handling arrival flights is more highly associated with a general ability than either overflights or departures but equally associated with a general PS ability.

Ability determinants of TRACON test conditions. Correlations between the General ability and general PS ability composites and TRACON performance, overall and by flight type, are shown in Figures 4 and 5. If the ability–performance correlations tracked the changes in mean performance levels, we would anticipate that the correlations between general ability and performance in the overflight-only conditions would be lower than during training, and the correlations with general ability would be higher in arrival conditions—especially so in the arrivals–65% condition. In contrast, we predicted that only the content manipulations (i.e., the contrast between arrival and overflight test conditions) would result in significant differences in ability–performance correlations.

In support of one part of our predictions, and shown clearly in Figure 4, there were no marked differences in correlations with either the arrivals or overflight flight types. A slight, but nonsignificant, increase in general ability correlations was observed from the 100% compliance to the 65% compliance conditions for the arrival condition, $t(78) = 1.02$, *ns*, and a negligible difference in correlations was found for the overflight condition, $t(78) = -0.11$, *ns*. Because performance in the respective conditions was highly correlated ($r = .829$ for the two overflight test conditions, and $r = .94$ for the two arrival conditions), the power to detect even small differences in correlations was quite high. Although exact power

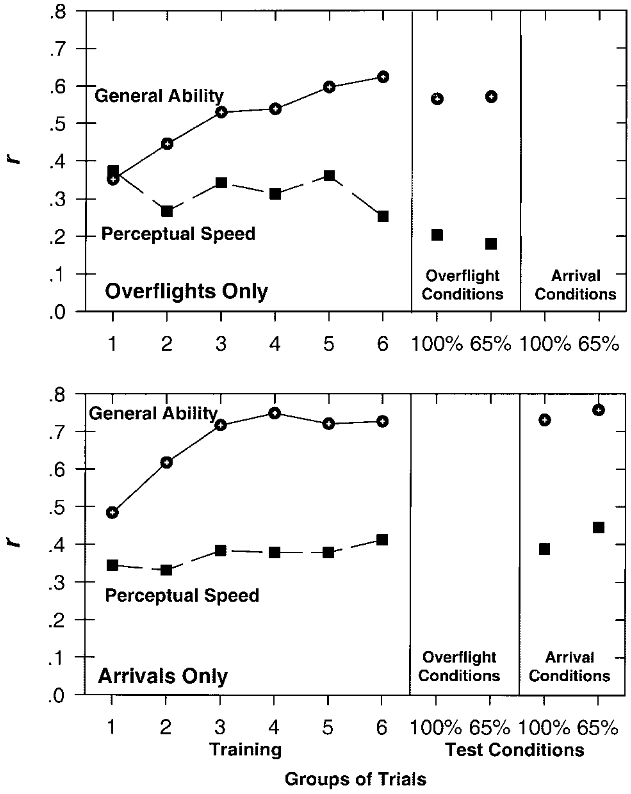


Figure 4. Correlations between General and Perceptual Speed ability composites and Terminal Radar Approach Control performance during training and test conditions. Test conditions were administered in counterbalanced order. Dark circles = General Ability; solid squares = Perceptual Speed.

cannot be computed, an example should suffice. With $df = 78$, a difference in correlations of ± .10 (from $r_{12} = .6$ to $r_{13} = .7$), where the two conditions are highly correlated ($r_{23} = .80$), is significant beyond the .05 level. As the correlation between conditions increases, any differences between their respective correlations with a third variable becomes easier to detect (see Ferguson, 1976, for the appropriate formula). Thus, we can confidently conclude that the reduced compliance condition had no meaningful effect on the ability determinants of task performance in either the arrival or overflight conditions.

In contrast to the compliance conditions, comparison of the content manipulations (arrivals vs. overflights) did result in significant differences between ability–performance correlations, which is a generalization of previous data from within-task comparisons of the two different types of flights (see Ackerman et al., 1995). That is, the average correlation between the general ability composite and arrivals-only performance was .744 but was .568 for overflights-only performance, which is a significant difference, $t(79) = 2.72$, $p < .01$. Similarly, the arrivals-only performance was more highly correlated than the overflights-only performance with the general PS ability (r = .417 and .192, respectively), $t(79) = 2.62$, $p < .01$. Therefore, our hypothesis that task content has a significant effect on ability–performance correlations was supported—handling arrival flights was more highly correlated with general and PS abilities than handling overflights.

Examination of the number of departures handled within each of the four test conditions, shown in the top panel of Figure 5, revealed no significant differences between respective correlations with the general ability composite (average $r = .560$ for arrival conditions and .507 for overflights), $t(79) = -0.93$, *ns*, or the PS ability composite (average $r = .429$ for arrival conditions and .334 for overflight conditions), $t(79) = 1.53$, *ns*, though the latter difference just failed to reach significance.

Regardless of the lack of differences among the separate flight types within conditions, however, it is important to note that the general ability and the general PS ability are significantly related to both individual flight types and overall performance (shown in the bottom panel of Figure 5). Indeed, the general ability composite correlated an average of .697 with overall performance in the four test conditions, and the general PS composite correlated an average of .382 with overall performance in the four test conditions. Because of the common variance between the general and PS composites, though, the aggregate predictive validity of the two composites is not much larger than the respective raw correlations with the general composite. The range of predictive validities was from $R^2 = .43$ to .56 for the four test conditions, indicating that these abilities account for 43%–56% of the individual-differences variance in task performance. It is interesting to also note that the general PS composite showed significantly different correlations with aggregate performance in the overflight versus arrival conditions (see Figure 5, lower panel). The mean correlation between the general PS ability and overall performance in the overflight conditions was .297, whereas the mean correlation between PS and arrival conditions was significantly larger ($r = .461$), $t(78) = 2.60$, $p < .01$.

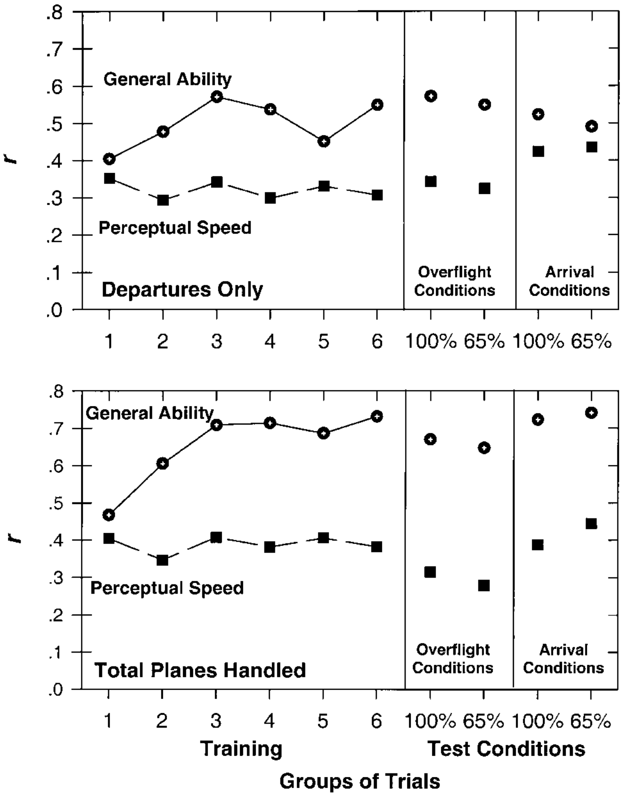


Figure 5. Correlations between General and Perceptual Speed ability composites and Terminal Radar Approach Control performance during training and test conditions. Test conditions were administered in counterbalanced order. Dark circles = General Ability; solid squares = Perceptual Speed.

Discussion and Conclusions

There is considerable complexity that arises when one attempts to integrate the two disciplines of scientific psychology (Cronbach, 1957). In the case of the current study, we examined a complex simulation task in the context of experimental psychology (e.g., means and overall treatment effects) and the context of correlational psychology (e.g., the relations between abilities and performance under different experimental conditions). However, this was not designed as an examination of aptitude–treatment interactions, because our goal was not to maximize criterion task performance. Rather, we were interested in whether substantial experimentally induced performance effects are in some fashion dissociated from ability–performance correlations.

The experimental effects were generally quite tractable, as expected. Increasing information-processing demands by increasing the number of arrival flights and eliminating overflights resulted in decreased performance. Conversely, eliminating arrival flights and increasing overflights resulted in increased performance. Reducing the consistency between participant inputs and system outputs by reducing pilot compliance from 100% to 65% also reduced overall performance, especially in the task with the highest initial information-processing demands (i.e., the arrivals condition). Subjective task difficulty was largely in agreement with the overall performance data. The participants found the reduced compliance conditions to be more difficult than the perfect compliance conditions, the arrival conditions to be more difficult than the overflight conditions, and the arrival–65% compliance condition to be of much greater difficulty than the others.

The central issue about the isomorphism between experimental manipulation effects on performance means and ability–performance correlations provides a more complex perspective. Changes in compliance levels, which was a change in the consistency of participant input and system output, although having substantial effects on task performance, failed to significantly affect the relationships between general and PS ability on the one hand and individual differences in performance on the other hand. However, changes in the content–complexity of the task (arrivals vs. overflights) were reflected in ability–performance correlations. Having a preponderance of flight types (arrivals) that had greater spatial and perceptual-speed demands yielded comparatively higher correlations between performance and the respective ability composites. Conversely, having a preponderance of flight types (overflights) that had reduced spatial and perceptual-speed demands yielded comparatively lower correlations between performance and the respective ability composites. Thus, only some task characteristic manipulations that affected mean performance also affected ability–performance relationships.

For experimentally oriented applied differential psychologists, the robustness of ability–performance relationships found in the larger industrial and educational selection literature has been an especially vexing problem. On the one hand, numerous investigators, such as Ree and his colleagues (see e.g., Ree, Earles, & Teachout, 1994) and Hunter (1986), have suggested that measures of general intelligence are effective "universal" predictors of job performance. Although one may easily object to aspects of the data sets collected by these investigators (e.g., neither the Armed Forces Vocational Aptitude Battery nor the General Ability Test Battery, which figure prominently in meta-analyses of validity generalization studies, is generally adequate for assessing multiple different abilities), a substantial number of datasets has been accumulated that fail to show differential patterns of ability–performance relations for different jobs or tasks. There are relatively few investigations that, in contrast, show that tailored selection tests can yield more effective selection predictions than measures of general intelligence (though see, e.g., Ackerman & Kanfer, 1993, for one example of the selection of air traffic controllers; and see Wittmann & Süß, 1999, for other empirical examples). However, there has been some indication that task complexity and/or difficulty do change the overall level of ability–performance correlations (see Hunter, 1986).

The results of the current study are relatively preliminary in terms of potential generalizability to real-world selection. However, the main implication of these results is that only some task characteristics, as can be determined from cognitive task analysis, are important determining factors for ability–performance relations. The most salient factor in the current investigation is that of task content–complexity. Consistent with earlier results with basic information-processing tasks (Ackerman, 1986; see also Ackerman & Kyllonen, 1991; Kyllonen, 1985), the content of the task, whether it be spatial, verbal, or numerical, along with differentiable demands on processing speed or perceptual speed, appears to be much more influential than the consistency of participant input and system output. We expect that classification of tasks and jobs by the content dimension may be much more useful for development of more effective selection instruments than either a focus only on general ability or on a task decomposition that focuses, for example, on the stages of information processing required by different tasks. Selection batteries that focus on content (e.g., verbal tests for jobs requiring processing of verbal materials, spatial tests for jobs requiring processing of spatial materials, and so on) may turn out to have much greater selection utility than either basic information-processing inspired psychometric measures or broad general reasoning tests.

References

Ackerman, P. L. (1986). Individual differences in information processing: An investigation of intellectual abilities and task performance during practice. *Intelligence, 10,* 101–139.

Ackerman, P. L. (1987). Individual differences in skill learning: An integration of psychometric and information-processing perspectives. *Psychological Bulletin, 102,* 3–27.

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of Experimental Psychology: General, 117,* 288–318.

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

Ackerman, P. L. (2000). A reappraisal of the ability determinants of individual differences in skilled performance. *Psychologische Beiträge, 42,* 4–17.

Ackerman, P. L., Bowen, K. R., Beier, M. B., & Kanfer, R. (2001). Determinants of individual differences and gender differences in knowledge. *Journal of Educational Psychology, 93,* 797–825.

Ackerman, P. L., & Cianciolo, A. T. (2000). Cognitive, perceptual speed, and psychomotor determinants of individual differences during skill acquisition. *Journal of Experimental Psychology: Applied, 6,* 259–290.

Ackerman, P. L., & Kanfer, R. (1993). Integrating laboratory and field study for improving selection: Development of a battery for predicting air traffic controller success. *Journal of Applied Psychology, 78,* 413–432.

Ackerman, P. L., Kanfer, R., & Goff, M. (1995). Cognitive and noncognitive determinants and consequences of complex skill acquisition. *Journal of Experimental Psychology: Applied, 1,* 270–304.

Ackerman, P. L., & Kyllonen, P. C. (1991). Trainee characteristics. In J. E. Morrison (Ed.), *Training for performance: Principles of applied human learning* (pp. 193–229). West Sussex, England: Wiley.

Ackerman, P. L., & Woltz, D. J. (1994). Determinants of learning and performance in an associative memory/substitution task: Task constraints, individual differences, and volition. *Journal of Educational Psychology, 86,* 487–515.

Andrew, D. M., & Paterson, D. G. (1934). Measured characteristics of clerical workers. *Bulletins of the Employment Stabilization Research Institute, University of Minnesota, 3 (1),* 7–19.

Barrett, G. V., Alexander, R. A., & Doverspike, D. (1992). The implications for personnel selection of apparent declines in predictive validities over time: A critique of Hulin, Henry, and Noon. *Personnel Psychology, 45,* 601–617.

Bennett, G. K., & Gelink, M. (1951). *CA-2.* New York: Psychological Corporation.

Buckley, E. P., DeBaryshe, B. D., Hitchner, N., & Kohn, P. (1983). *Methods and measurements in real-time air traffic control system simulation* (Rep. No. DOT/FAA/CT-83/26). Atlantic City Airport, NJ: U.S. Department of Transportation, Federal Aviation Administration.

Burke, H. R. (1958). Raven's Progressive Matrices: A review and critical evaluation. *Journal of Genetic Psychology, 93,* 199–228.

Carroll, J. B. (1980). *Individual difference relations in psychometric and experimental cognitive tasks* (Tech. Rep. No. 163). Chapel Hill: University of North Carolina, The L. L. Thurstone Psychometric Laboratory.

Carroll, J. B. (1989). *Factor analysis programs.* Chapel Hill, NC: Author.

Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies.* New York: Cambridge University Press.

Cobb, B. B., & Mathews, J. J. (1972). *A proposed new test for aptitude screening of air traffic controller applicants* (FAA-AM-72-18). Washington, DC: U.S. Department of Transportation, Federal Aviation Administration.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Erlbaum.

Cronbach, L. J. (1957). The two disciplines of scientific psychology. *American Psychologist, 12,* 671–684.

Deary, I. J., & Stough, C. (1996). Intelligence and inspection time: Achievements, prospects, and problems. *American Psychologist, 51,* 599–608.

Ferguson, G. A. (1976). *Statistical analysis in psychology and educations* (4th ed.). New York: McGraw-Hill.

Guilford, J. P., & Lacey, J. I. (Eds.). (1947). *Army air forces aviation psychology program research reports: Printed classification tests* (Rep. No. 5). Washington, DC: U.S. Government Printing Office.

Hulin, C. L., Henry, R. A., & Noon, S. L. (1990). Adding a dimension: Time as a factor in the generalizability of predictive relationships. *Psychological Bulletin, 107,* 1–13.

Humphreys, L. G., & Montanelli, R. G., Jr. (1975). An investigation of the parallel analysis criterion for determining the number of common factors. *Multivariate Behavioral Research, 10,* 193–206.

Hunt, E., Frost, N., & Lunneborg, C. (1973). Individual differences in cognition: A new approach to intelligence. In G. Bower (Ed.), *Advances in learning and motivation, Volume 7* (pp. 87–122). New York: Academic Press.

Hunter, J. E. (1986). Cognitive ability, cognitive aptitudes, job knowledge, and job performance. *Journal of Vocational Behavior, 29,* 340–362.

Jensen, A. R. (1998). *The g factor: The science of mental ability.* Westport, CT: Praeger.

Keeling, K. B. (2000). A regression equation for determining the dimensionality of data. *Multivariate Behavioral Research, 35,* 457–468.

Kyllonen, P. C. (1985). *Dimensions of information processing speed* (Rep. No. AFHRL-TP-84-56). Brooks Air Force Base, TX: Air Force Systems Command.

Lachman, R., Lachman, J. L., & Butterfield, E. C. (1979). Cognitive psychology and information processing: An introduction. Hillsdale, NJ: Erlbaum.

Ree, M. J., Earles, J. A., & Teachout, M. S. (1994). Predicting job performance: Not much more than *g. Journal of Applied Psychology, 79,* 518–524.

Rosenthal, R., Rosnow, R. L., & Rubin, D. B. (2000). *Contrasts and effect sizes in behavioral research: A correlational approach.* New York: Cambridge University Press.

Shiffrin, R. M., & Schneider, W. (1977). Controlled and automatic human information processing: II. Perceptual learning, automatic attending, and a general theory. *Psychological Review, 84,* 127–190.

Spearman, C. E. (1938). Measurement of intelligence. *Scientia Milano, 64,* 75–82.

Sternberg, R. J. (1977). *Intelligence, information processing, and analogical reasoning: The componential analysis of human abilities.* Hillsdale, NJ: Erlbaum.

Thorndike, R. L. (1986). The role of general ability in prediction. *Journal of Vocational Behavior, 29,* 332–339.

Tucker, L. R., & Finkbeiner, C. T. (1981). *Transformation of factors by artificial personal probability functions* (Rep. No. RR-81-58). Princeton, NJ: Educational Testing Service.

Wechsler, D. (1958). *The measurement and appraisal of adult intelligence* (3rd ed.). Baltimore: Williams & Wilkins.

Wittmann, W. W., & Süß, H.-M. (1999). Investigating the paths between working memory, intelligence, knowledge, and complex problem-solving performances via Brunswik Symmetry. In P. L. Ackerman, P. C. Kyllonen, & R. D. Roberts (Eds.), *Learning and individual differences: Process, trait, and content determinants* (pp. 77–108). Washington, DC: American Psychological Association.

Woltz, D. J. (1988). An investigation of the role of working memory in procedural skill acquisition. *Journal of Experimental Psychology: General, 117,* 319–331.

Received February 9, 2002
Revision received March 25, 2002
Accepted April 12, 2002 ■