Psychology and Aging
1990, Vol. 5, No. 3, 437–446

0882-7974/90/$00.75

Structural Equation Models of Relationships Between Exercise and Cognitive Abilities

Louise Clarkson-Smith and Alan A. Hartley
Scripps College

Data were obtained from 300 men and women aged 55 to 91. Separate structural equation models of relationships between physical exercise and 3 cognitive performance variables—reaction time, working memory, and reasoning—fit the data well. Other variables in the models were age, health, education, and morale. Age and exercise affected each performance variable directly; education had a direct effect on reasoning only. There were also indirect effects of age and health on performance variables, mediated through exercise. The main hypothesis of the study, that exercise contributes to performance, was supported. A large decrease in model fit resulted when the path from exercise to each performance variable was deleted. Hypotheses that age-related deficits are primarily accounted for by lack of exercise or by poor health were not supported.

Recent research suggests that, for older adults, there is a positive relationship between physical exercise and performance on various cognitive tasks. Correlational studies, in which the performance of habitual exercisers is compared with that of sedentary individuals, have reported superior performance for older exercisers on reaction time (RT) tasks (Baylor & Spirduso, 1988; Clarkson-Smith & Hartley, 1989; Spirduso, 1980), nonverbal reasoning tasks (Clarkson-Smith & Hartley, 1989; Elsayad, Ismail, & Young, 1980; Powell & Pohndorf, 1971), and working memory tasks (Abourezk, 1988; Clarkson-Smith & Hartley, 1989; Ohlsson, 1975). Although correlational studies have consistently shown a positive relationship between physical exercise and cognitive performance, they lack the experimental control necessary to establish a causal relationship.

Results are less consistent from intervention studies that compare the performance of participants randomly selected to be in either an exercise training program or a control group. No improvement in a visual search task was reported as a function of aerobic training by Blumenthal and Madden (1988) for a middle-aged sample or by Madden, Blumenthal, Allen, and Emery (1989) for an older sample. In both studies, however, the aerobic group improved in aerobic capacity. Blumenthal and Madden also reported that faster performance at the initial testing was associated with greater aerobic capacity. The performance of an older aerobic-exercise group of Dustman et al. (1984) improved in aerobic capacity, as well as in several cognitive measures, but no improvement as a function of exercise was seen for nonverbal reasoning or choice RT tasks. Surprisingly, scores on nonverbal reasoning tests for the control group showed an improvement at the end of the 4-month training period, suggesting that effects of exercise may be confounded with practice effects in a short-term experimental study.

Thus, no studies have been able to establish a causal relationship between exercise and complex processes such as reasoning in older people. Although intervention studies have the advantage of the experimental control necessary to infer causality, correlational studies have the advantage of greater ecological validity in that the participants may be individuals for whom exercise is a well-established part of their life-style. In this study, we attempted to further investigate the relationship between habitual exercise and cognitive abilities in older populations by using a subvariety of the correlational approach, structural equation modeling (Bentler, 1980; Hayduk, 1987; Jöreskog & Sörbom, 1984), which is a means of testing theory-based hypotheses with nonexperimental data.

In formulating a model of the postulated effect of exercise on cognition in older people, we included variables that have been related to both age and cognition. Steuer and Jarvik (1981) suggested that age-related cognitive deficits may result from the cumulative effects of health, education, and aging. These are variables that may also affect exercise. Therefore, exercise may serve as a mediating variable between each of these variables and cognitive performance. For example, there is strong empirical support for a negative relationship between age and performance on a large number of cognitive tasks (Charness, 1985; Horn, 1980; Reese & Rodeheaver, 1985), but there is also consistent evidence that people become less active with advancing age (Ostrow, 1984; Stones & Kozma, 1985). Thus, the decrease in physical activity of older people might account for a portion of age-related deficits in cognition.

The direction of causal influence among the variables

The research was supported in part by National Institute on Aging Postdoctoral Fellowship Grant AG05370 to Louise Clarkson-Smith.

We thank Amy Gamble for technical assistance and Christopher Hertzog and Keith Widaman for their extremely helpful suggestions on a draft of this article. We are grateful for the interest and cooperation shown by the participants in the study and for the assistance of the following organizations and publications in locating participants: Claremont Adult School, The Claremont Club, Claremont University Club, Corona del Mar Track Club, Hillcrest Homes, Joslyn Center, *Masters Journal*, Peer Counselors, Recording for the Blind, San Dimas Women's Club, Saint Ambrose Church, and the Service Corps of Retired Executives.

Correspondence concerning this article should be addressed to Louise Clarkson-Smith, Scripps College, 1030 Columbia Avenue, Claremont, California 91711-3948.

(health, physical exercise, and cognition) is not clear. Research suggests that exercise has a beneficial effect on various aspects of health (Bortz, 1982; Cunningham, Rechnitzer, Howard, & Donner, 1987; deVries, 1983). Studies investigating relationships of cognitive performance to cardiovascular disease (Hertzog, Schaie, & Gribbin, 1978) and blood pressure (Schultz, Elias, Robbins, Streeten, & Blakeman, 1986) suggest that cognition is associated with cardiovascular fitness, the aspect of health that is particularly enhanced with exercise. Therefore, the effect of exercise may be mediated through one aspect of health, cardiovascular fitness. At the same time, various forms of poor health may have an indirect negative effect on cognition by limiting exercise. The possibility that physical exercise has cognitive benefits that are independent of health is suggested by the results of Clarkson-Smith and Hartley (1989). In a comparison of older exercisers and nonexercisers, they found that self-rated health did not account for a significant proportion of the variance in cognitive performance and that the performance of a group of exceptionally healthy exercisers was superior to the performance of a group of exceptionally healthy nonexercisers.

Although a positive relationship between education and performance on various cognitive tasks is well established, Milligan, Powell, Harley, and Furchtgott (1984) did not find a correlation between education and RT in older men. We are not aware of any research relating educational levels to physical exercise. Yet, individuals who are better educated may be better informed on the health value of exercise and consequently make more of an effort to engage in physically active pursuits.

Subjective well-being was included in the model because of its potential relation to both exercise and cognition. Manton, Siegler, and Woodbury (1986) found that superior performers on various cognitive tasks were also superior in measures of mental and emotional adjustment. On the other hand, Milligan et al. (1984) found that the positive correlation of measures of subjective well-being with RT and serial learning scores disappeared when health and education were partialed out. There are correlational data relating subjective well-being to exercise (see Ostrow, 1984, and Stones & Kozma, 1985, for reviews), but the few experimental studies are almost equally divided between those that report an improvement as a result of an exercise training program and those that report no change (Doan & Scherman, 1987).

To summarize, the proposed model included four variables believed to have a potential effect on both exercise and cognitive performance: age, health, education, and subjective well-being. Under the assumption that exercise might help to forestall age-related degenerative changes in the brain, we selected three classes of cognitive performance variables for which current theories (Craik & Byrd, 1982; Hasher & Zacks, 1979; Horn, 1980; Salthouse, 1985) suggest the age-related decline might have a biological rather than an experiential basis. The three classes of performance variables, ranging from simple to complex, were RT, working memory, and reasoning.

In the proposed model (Figure 1), age, education, health, and subjective well-being affect cognitive performance indirectly through exercise. In addition, there are direct causal paths from age to each of the performance variables and from education to reasoning.

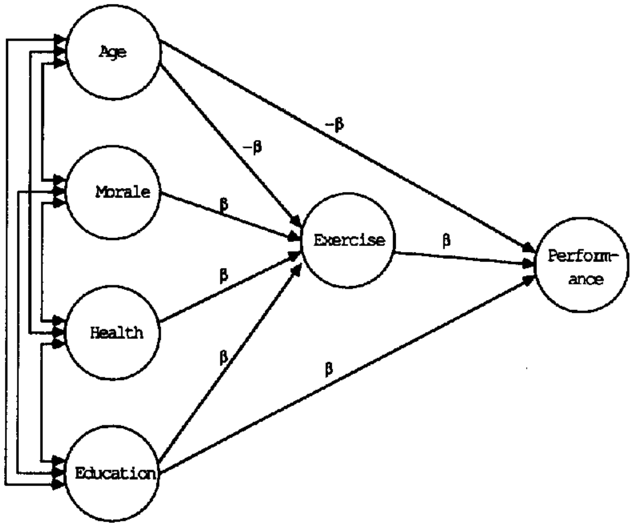

Figure 1. Model A: General model of exercise–performance relationships for reasoning and working memory. (The path from education to performance applied only to the model of reasoning. Paths from morale and education to exercise were not significant. For model of reaction time, the path from age to performance was positive, and the path from exercise to performance was negative. Single-headed arrows indicate hypothesized unidirectional influences; double-headed arrows indicate correlations with no specified direction of influence.)

Structural equation modeling has several advantages that make it particularly appropriate for this study. First, the method makes it possible to develop and test models, such as the proposed model, that contain complex interrelationships among variables, including direct effects of variables on each other and indirect effects mediated through intervening variables. A second advantage is that by using latent variables, estimates can be made of causal relationships among latent variables that are free of the biasing effects of error variance.

Particularly important to this study is the ability to test hypotheses using nested models, in which one model differs from another only in the addition or deletion of the hypothesized relationship. Three key hypotheses were tested in this manner. The first hypothesis was that physical exercise had a significant effect on cognitive performance in addition to the hypothesized effects of age, health, morale, and education. The second hypothesis was a stronger version of the first, that cognitive changes associated with aging are primarily the result of a sedentary life-style rather than of the aging process itself. In support of this hypothesis, Bortz (1982) has cited numerous examples of physical and mental correlates of aging that are similar to the effects of enforced bed rest in young adults. The third hypothesis, suggested as one possibility by Steuer and Jarvik (1981), holds that all or most cognitive decline with age is a function of declining health.

2

Method

1-1

Participants

Participants were 87 men and 213 women ranging in age from 55 to 91. Participants were recruited from community organizations and retirement communities. The sample was recruited for several studies investigating relations between cognition and various activities. Therefore, rather than attempting to obtain a representative sample, we gave the highest priority to recruiting individuals participating in various activities, including physical exercise. Participants were all living independently in the community or in retirement communities. They were screened for central nervous system disorders, including a history of strokes and transient ischemic attacks, and for visual disorders sufficiently severe to interfere with performance. Six potential participants were excluded because of history of stroke or other neurological disorders. The native language of all participants was English. One woman was Black, and 1 woman was of Asian descent. All other participants were White.

Procedure

Participants were tested in two sessions of approximately 1.5 hr each. The first session consisted of tests of working memory and RT and an interview to obtain detailed information on the amount and type of physical activity in which participants had engaged during the past year. Data on measures of physical fitness, resting heart rate, vital capacity, and blood pressure were also obtained at this session. The second session consisted of three written tests of reasoning and two questionnaires assessing subjective well-being. At home, participants completed a questionnaire covering age, health, education, and activities in which they were participating.

The following is a list of the observed variables included in one or more of the models:

1. *Age.* Age was self-reported chronological age, calculated in months from the date of birth.

2. *Self-rated health.* Health was rated on a 7-point scale (1 = *very poor,* 2 = *poor,* 3 = *fair,* 4 = *average,* 5 = *good,* 6 = *very good,* and 7 = *excellent*).

3. *Illnesses.* Participants were asked to list any medical conditions and medications they were taking. Conditions with a potential effect on cognition—cardiovascular disease, other heart ailments, high blood pressure (not corrected by medication), respiratory disorders, and diabetes—were scored 2. All other medical conditions were scored 1.

4. *Medications.* Medications listed in the *Physician's Desk Reference* (Barnhart, 1986) as possibly causing psychomotor slowing, short-term-memory impairment, or mental confusion were scored 2; all others were scored 1. To indicate good rather than poor health, scores for illnesses and medications were subtracted from 10 (9 was the highest number reported).

5. *Life Satisfaction Index A* (Neugarten, Havighurst, & Tobin, 1961). The Life Satisfaction Index is a test of subjective well-being designed for older people. It consists of 20 attitude statements (i.e., "I am just as happy as when I was younger"). Participants were to indicate agreement, disagreement, or uncertainty for each statement. The test was scored according to a system suggested by Wood, Wylie, and Sheafer (1969). Responses indicating a high level of satisfaction were scored 2, those indicating a low level of satisfaction were scored 0, and uncertain or uncommitted responses were scored 1.

6. *Feelings Scale* (from the Questionnaire for the Study of Modern Living; Bradburn & Caplovitz, 1965). The participants were to indicate the frequency that each of 12 feelings (e.g., bored, on top of the world) were experienced during the past week. Standard scoring was used, with positive scores for positive feelings and negative scores for negative feelings.

7. *Years of education.* We counted full-time or equivalent in part-time education. Preschool and kindergarten were excluded.

8. *Highest degree earned.* Degrees were scored as follows: 1 = *no degree or diploma,* 2 = *high school diploma,* 3 = *AA or nursing degree,* 4 = *BA or BS,* 5 = *MA or MS,* 6 = *law degree,* 7 = *PhD or MD.*

9. *Kilocalories per week.* Information on the amount and type of exercise obtained from interviews with participants was converted to kilocalories (Heyward, 1984). All activities were included that required physical exertion beyond that of normal activities of daily living, including not only athletic pursuits but also nonrecreational activities, such as heavy housework, gardening, walking, and climbing stairs.

10. *Strenuous exercise.* The estimated average number of hours per week of participation in exercise greater than 0.1 $\mathrm{kcal \cdot min \cdot kg^{-1}}$ of body weight (the equivalent of a very fast walk or a slow jog) was obtained from an interview.

11. *Previous exercise.* Participants were asked to compare, on a 5-point scale, their level of physical activity during the past 10 years with their current level of activity. Estimates were defined as percentages of the current number of kilocalories per week: 50%, 75%, 100%, 125%, and 150%, for *much less, somewhat less, the same, somewhat more,* and *a lot more,* respectively.

12. *Vital capacity.* The percentage of normal vital capacity was calculated from forced expiratory volume obtained with a Propper compact spirometer; this was adjusted for age, height, and gender. One participant refused to take this test.

13. *Heart rate.* Heart rate and blood pressure were obtained at the beginning of the first session with participants in a seated position. Heart rate and blood pressure were recorded from the left arm with a Sharp digital blood pressure monitor (Model MB-500).

14. *Blood pressure.* Systolic and diastolic blood pressures were measured. Because low, rather than high, scores on blood pressure and heart rate were considered measures of fitness, scores were converted to negative values.

15. *Reaction time.* Three types of RT were measured. RT tasks were presented on an Apple Macintosh computer. For simple RT (RT1), the stimulus was a large O presented in the center of the screen. Participants were instructed to rest the fingers of their preferred hand on the space bar and to press it as rapidly as possible after seeing the stimulus. For two-choice (RT2) and four-choice (RT4) tasks, the stimulus was a single digit: 1, 2, 3, or 4. The digits always appeared in the same location on the screen. The labels *1, 2, 3,* and *4* were applied to keys *D, F, J,* and *K.* For RT2, participants were instructed to rest their middle fingers on the keys labeled *1* and *4,* and for RT4, also to rest their index fingers on the keys labeled *2* and *3.* Participants were instructed to press the key corresponding to the digit that appeared on the screen as rapidly as possible, but to try to do it correctly. If subjects made an error, they were to ignore it and continue with the next response as rapidly as possible. Each stimulus was preceded by a 0.32-s tone followed by a random variable interval ranging from 0.1 to 3 s. The stimulus remained on the screen until a key was pressed. After 2 s, the tone preceding the next stimulus was sounded. There were five practice trials for RT1 and 10 practice trials for RT2 and RT4. The value for each of the RT measures was the geometric mean of response latencies for 30 correct trials.

16. *Letter sets.* This task, which was the first of three measures of memory span, was a modification of a task introduced by Crawford and Stankov (1983). Two sets of two, three, or four letters were presented sequentially on a screen with a 1.5-s interval between the offset of the first stimulus and the onset of the second stimulus. One letter in each set was different; the others were the same (e.g., *LFG* and *GLS*). Participants were instructed to read the letters aloud as they appeared on the screen and after the second set disappeared, to name the letter that was unique to each set. Initially, sets of two letters were presented for 6 trials, followed by 10 trials of three-letter sets and 10 trials of four-letter sets. Letters appeared on the screen just long enough for them to be read aloud, 1.5 s for two letters, 2.0 s for three letters, and 2.5

s for four letters. Responses were not timed. Three practice trials were presented at the beginning of the exercise, before starting the three-letter sets and before starting the four-letter sets. The exercise was terminated after five consecutive errors. The score was the number of correct responses, weighted by the number of letters in each set for which a correct response was given, and summed across sets.

17. *Digit-span backward* (from the Wechsler Adult Intelligence Scale; Wechsler, 1955). To be comparable with scoring of the letter-sets task, the score was the number of correct responses weighted by the number of digits in each set and summed across sets. Both digit-span-forward and digit-span-backward tasks were given, but only digit-span backward was used in the analysis.

18. *Reading span.* The task was adapted from one by Daneman and Carpenter (1980), in which the participant reads aloud a set of sentences and is then asked to recall the last word of each sentence. In this study, set size ranged from two to six, with three sets of sentences at each level. The task was discontinued after the participant failed to make at least one correct response at any one level. Three practice trials were given at the beginning of the task. The score was the number of correct responses weighted by the number of sentences.

19. *Analogies.* The first of the three indicators of reasoning was a task consisting of 30 common-word verbal analogies graded for difficulty (Clarkson-Smith, 1985). Analogies were presented in a forced-choice format with three answer options. All three reasoning tasks were paper-and-pencil tests and were untimed.

20. *Matrices.* Twenty graded items were selected from Advanced Progressive Matrices, Sets I and II (Raven, 1972, 1974).

21. *Letter series.* There were 15 letter-series-completion items (Horn, 1975). The participant was to name, on the basis of the ordering of letters in the alphabet, the next letter in a series of letters.

The results for each of the above variables for the total sample and for men and women are shown in Table 1. There were consistent differences in favor of men for exercise and education. Men were also superior in one RT task and two reasoning tasks, whereas women were superior in one test of morale.

Results

Separate models were tested for each of the three performance variables. A variance–covariance matrix of the variables just listed was the input data. In all analyses, parameter estimates were obtained for the models with the maximum likelihood option of LISREL VI (Jöreskog & Sörbom, 1984). Goodness of fit of the models was evaluated with the chi-square statistic, a measure of the difference between the predicted and observed covariance matrices. The chi-square was used to test hypotheses by assessing the difference between nested models, models that differ only in the inclusion of one or more additional parameters. Because chi-square is strongly influenced by sample size, we also used a measure of practical fit, the Tucker and Lewis rho coefficient (Bentler & Bonett, 1980), which is less affected by sample size than is the chi-square (Marsh, Balla, & McDonald, 1988, but see also Bollen, 1986) and is considered a more reliable index of fit when assumptions of multivariate normality are not met (Bentler, 1988). According to Bentler and Bonett, an adequately fitting model should have a rho value of at least .9.

Following the recommendation of Anderson and Gerbing (1988), we first estimated the measurement model, which specifies relationships of the observed variables to the underly-

Table 1
Values of Observed Variables

	Entire sample		Men		Women		
Variable	*M*	*SD*	*M*	*SD*	*M*	*SD*	*t*
Age	70.13	7.22	71.29	6.99	69.65	7.27	1.82
Health	5.65	1.09	5.71	1.13	5.62	1.07	0.62
Illnesses	9.06	1.25	9.01	1.27	9.04	1.24	0.41
Medications	8.67	1.80	8.97	1.58	8.55	1.87	1.96
Life satisfaction	32.59	4.89	32.44	4.85	32.65	4.91	−0.34
Feeling Scale	5.71	3.85	4.98	3.92	6.00	3.78	−2.08*
Education	15.95	2.45	16.52	2.67	15.72	2.33	2.44*
Degree	3.74	1.43	4.07	1.53	3.60	1.37	2.48*
Kilocalories	2,351.77	1,578.27	3,319.67	1,868.23	1,956.43	1,247.32	6.26**
Hours of exercise	2.54	3.01	3.25	3.41	2.25	2.79	2.42*
Previous exercise	2,386.42	1,476.04	3,211.52	1,646.19	2,047.83	1,255.62	5.92**
Vital capacity	0.82	0.19	0.83	0.18	0.82	0.19	0.56
Heart rate	68.80	10.44	67.39	10.33	69.38	10.46	−0.33
Systolic BP	133.45	17.70	131.40	18.88	134.29	17.17	−1.23
Diastolic BP	78.01	9.85	77.71	10.04	78.13	9.78	0.67
RT1	412.89	53.56	397.92	39.46	419.00	57.31	3.65**
RT2	522.12	77.69	511.78	84.25	526.35	74.65	1.40
RT4	722.59	117.13	704.93	109.34	729.81	119.67	1.74
Letter sets	40.83	15.39	40.98	15.86	40.77	15.23	0.10
Digit-span backward	25.77	14.61	24.85	14.53	26.14	14.66	−0.70
Reading span	13.17	7.67	12.72	6.05	13.36	8.24	−0.74
Analogies	22.42	4.33	23.17	3.84	22.11	4.49	2.07*
Matrices	13.77	3.80	14.11	3.60	13.63	3.88	1.04
Series	11.18	3.06	11.72	2.62	10.95	3.20	2.16*

Note. BP = blood pressure; RT = reaction time.
* $p < .05$. ** $p < .001$.

ing constructs or latent variables, and then the structural model, which specifies causal relationships among the latent variables.

Measurement Model

The initial measurement model consisted of nine latent variables: age, health, morale, education, exercise, fitness, speed, working memory, and reasoning. All correlations among the nine latent variables were freely estimated. The loading of age, a single indicator, was fixed at 1.0, because it was assumed to be measured without error. Illnesses, medications, and self-rated health defined the second latent variable, health. The third latent variable, morale (subjective well-being), was defined by the Life Satisfaction Index and the Feelings Scale. The fourth latent variable, education, was defined by years of education and highest degree achieved. Variables defining the latent variable exercise were the average number of kilocalories per week, average number of hours of strenuous exercise per week, and previous activity level. The latent variable fitness was defined by vital capacity, resting heart rate, and systolic and diastolic blood pressures. The variable RT was defined by RT1, RT2, and RT4. Indicators for working memory were letter sets, reading span, and digit-span backward; matrices, analogies, and series served as indicators for reasoning.

A measurement model consists of three matrices: λ_y, a matrix of factor loadings of the observed variables on the appropriate latent variables; ψ, a variance–covariance matrix of the latent variables; and θ_ϵ, a matrix of error variances and covariances of the observed variables. Following the recommendation of Anderson and Gerbing (1988), the diagonal of the ψ matrix was fixed at 1.0, with loadings of each of the observed variables on the λ_y matrix free to vary. Three a priori nonzero correlations of errors in θ_ϵ were specified, between (a) kilocalories per week and previous activity level, (b) diseases and medications, and (c) series and matrices. As described earlier, the value for previous activity level was a percentage of the current weekly expenditure of kilocalories. Thus, any error in the estimation of the weekly expenditure of kilocalories would necessarily be reflected in the estimate of previous activity level. Two estimates of health, diseases and medications, contained only limited information on the type of condition or medication and no information on the severity of the illness and on the amount of medication. Therefore, the same type of error should have occurred in both estimates. The third error correlation was between the two nonverbal reasoning tasks. We expected that performance in each task could be influenced by visuospatial ability, an ability that would not influence performance on the verbal analogy task.

The initial analysis failed to converge to a solution, apparently because of model misspecification. Several indicators—large standard errors, large normalized residuals, and failure of the observed variables to load significantly on the fitness variable (Anderson & Gerbing, 1988; Jöreskog & Sörbom, 1984)—suggested that the problem lay with the fitness variable. Because there was no other appropriate latent variable to which these variables might be assigned, they were removed from the model (Anderson & Gerbing, 1988). The final measurement model was an adequate fit to the data, $\chi^2(140, N = 300) = 185.36$, $p = .006$, $\rho = .976$. Standardized factor loadings, t values, and error variances of the measurement model are shown in Table 2; correlations among factors are shown in Table 3. Convergent validity was indicated by t values (the ratios of the parameter estimates to the standard errors), which were all greater than 2 (Anderson & Gerbing, 1988). Discriminant validity was indicated by the finding that the confidence interval surrounding each correlation estimate did not include 1 (Anderson & Gerbing, 1988).

Table 2
Measurement Model: Standardized Parameter Estimates

Factors and variables	Factor loading	t	Error variance
Age	1.000[a]	—	.000[a]
Health			
Self-rated health	.957	9.549	.083
Illnesses	.474	6.571	.784
Medications	.465	6.665	.755
Morale			
Life satisfaction	.726	6.706	.423
Feelings Scale	.490	5.792	.760
Education			
Years	.944	18.927	.109
Highest degree	.950	19.101	.097
Exercise			
Kilocalories/week	.717	11.404	.486
Strenuous exercise (hr)	.908	13.892	.175
Previous exercise (hr)	.534	8.469	.715
Reaction time			
RT1	.672	12.196	.548
RT2	.840	16.117	.295
RT4	.802	15.356	.357
Working memory			
Letter sets	.752	12.809	.434
Reading span	.622	10.425	.613
Digit-span backward	.527	8.606	.722
Reasoning			
Matrices	.825	16.075	.429
Analogies	.755	13.992	.319
Series	.776	14.521	.397

Note. Dash indicates there were no t values for fixed parameters.
[a] Fixed parameter.

Structural Models

Because of the complexity of the model, we chose to test separate models for each performance variable, that is, RT,

Table 3
Measurement Model: Correlations of Factors

Variable	1	2	3	4	5	6	7	8
1. Age	—							
2. Health	−.166*	—						
3. Morale	−.033	.383*	—					
4. Education	−.021	.191*	.192*	—				
5. Exercise	−.308*	.355*	.259*	.202*	—			
6. Reaction time	−.432*	.249*	.183*	.102	.332*	—		
7. Working memory	−.377*	.196*	.026	.132	.309*	.453*	—	
8. Reasoning	−.402*	.242*	.180*	.443*	.403*	.466*	.757*	—

* $p < .05$.

Table 4
Standardized Factor Loadings and Error Variances of Model A for Reaction Time, Working Memory, and Reasoning

Variable	Reaction time		Working memory		Reasoning	
	Loading	Error	Loading	Error	Loading	Error
Age	1.000[a]	.000[a]	1.000[a]	.000[a]	1.000[a]	.000[a]
Health						
Self-rated	.946	.104	.948	.101	.944	.110
Illnesses	.470	.774	.469	.780	.471	.778
Medications	.480	.769	.479	.770	.481	.768
Morale						
Life satisfaction	.724	.476	.721	.480	.730	.468
Feelings Scale	.492	.758	.492	.758	.486	.764
Education						
Years	1.000	.000[a]	1.000	.000[a]	.962	.074
Degree	.897	.196	.897	.196	.932	.132
Exercise						
Kilocalories	.729	.469	.698	.513	.720	.481
Hours	.892	.205	.933	.130	.903	.185
Previous	.543	.705	.519	.731	.536	.713
Reaction time						
RT1	.659	.565	—	—	—	—
RT2	.853	.272	—	—	—	—
RT4	.799	.361	—	—	—	—
Working memory						
Letter sets	—	—	.676	.543	—	—
Reading span	—	—	.709	.497	—	—
Digit-span backward	—	—	.527	.722	—	—
Reasoning						
Matrices	—	—	—	—	.737	.457
Analogies	—	—	—	—	.850	.277
Series	—	—	—	—	.752	.435

Note. Dashes represent values that were not estimated.
[a] Fixed parameter.

working memory, and reasoning. For the initial analyses of RT and working memory, a small, negative error variance appeared for education. Because the negative estimates were close to 0 and the standard errors were not unduly large, the negative value appeared to be due to sampling fluctuation (Dillon, Kumar, & Mulani, 1987). Thus, following the recommendation of Dillon et al., the negative estimate was fixed at 0.

The proposed model (Model A, shown in Figure 1), proved to be an excellent fit to the data for RT, $\chi^2(65, N = 300) = 67.55$, $p = .387$, $\rho = .997$; for working memory, $\chi^2(65, N = 300) = 77.60$, $p = .136$, $\rho = .989$; and for reasoning, $\chi^2(62, N = 300) = 65.58$, $p = .354$, $\rho = .997$. Parameter estimates for models of RT, working memory, and reasoning are shown in Tables 4 and 5. Table 4 gives standardized factor loadings and error variances of the observed variables. Standardized path coefficients are shown in Table 5. With the exception of a causal path from education to reasoning, the parameters of the three models were quite similar. All causal paths, except those from morale and education to exercise, were significant.

Before testing hypotheses, some preliminary analyses were performed to investigate possible biasing effects of gender differences in exercise and of skewed distributions of some variables. We tested a revised version of Model A in which gender was added as an exogenous variable. The path from gender to exercise was significant. However, the addition of causal paths from gender to the performance variables did not significantly increase the chi-square, nor were the added paths significant. Furthermore, for models of RT and reasoning there were no changes from Model A in the significance of the other causal paths. The path from exercise to working memory, however, dropped slightly below the level of significance, suggesting that the relationship of exercise to working memory may differ be-

Table 5
Standardized Parameter Estimates of Directed Paths of Best Fitting Model for Reaction Time, Working Memory, and Reasoning

Pathway	Reaction time	Working memory	Reasoning
Age to exercise	−.262*	−.253*	.316*
Health to exercise	.254*	.236*	.246*
Morale to exercise	.141	.140	.135
Education to exercise	.101	.112	.117
Age to performance	−.355*	−.329*	−.316*
Education to performance	—	—	.407*
Exercise to performance	.232*	.198*	.224*

Note. Dashes represent parameters that were not estimated.
* $p < .05$.

Table 6
Correlations of Exercise-Measured Variables With Working-Memory-Measured Variables for Men (n = 87) and Women (n = 213)

	Kilocalories per week		Hours of strenuous exercise		Previous exercise level	
Variable	Men	Women	Men	Women	Men	Women
Letter sets	.300**	.189*	.275**	.236***	.286**	.175**
Reading span	.135	.236***	.160	.187**	.117	.032
Digit-span backward	.070	.175*	.117	.132*	.102	.019

* $p < .05$. ** $p < .01$. *** $p < .001$.

tween men and women. To further investigate this relationship, we compared, for men and women, correlations of measured variables of exercise with those of working memory (Table 6). For both men and women, there were strong correlations of all three exercise variables with letter sets. There were gender differences in significance for the correlation of hours of strenuous exercise with digit-span backward and reading span, but these were largely due to differences in sample size. The results therefore do not suggest that there were substantial gender differences in the relationship of exercise to any of the performance variables.

Further analyses were performed by square root or logarithmic transformations of variables with skewed distributions. Significance levels of the analyses and the pattern of parameters for causal paths did not differ from those of the untransformed data. The one exception was that the path from education to exercise reached significance. Because there were only minimal differences resulting from the preliminary analyses, subsequent analyses were performed on Model A with untransformed data.

Hypothesis Testing

Competing models representing alternative hypotheses about the three performance variables were tested against Model A. Directional paths for competing hypotheses are shown in Figure 2. For clarity, nonsignificant paths from education and morale to exercise, which were retained in Model A, are not shown in Figure 2. Table 7 shows results for alternate models. Table 8 gives the indices of difference between models for the comparisons that were made.

Before testing competing hypotheses, we compared Model A with a fully saturated model (Msat), in which all paths from the four endogenous variables—age, health, education, and morale—to exercise and to the performance variables, and from exercise to the performance variables, were estimated. In the comparison of Msat with Model A (Table 8), the restrictions of Model A did not significantly reduce the fit of the models. As a result, the parsimony of Model A supports it as the better representation of the data for each performance variable.

It has been suggested that age-related deficits in cognition are primarily the result of the sedentary life-style of older adults rather than a natural consequence of aging (Bortz, 1982). To test this hypothesis (Model B), we eliminated the direct path from age to each performance variable (β_1), leaving only the indirect path mediated through exercise (β_2, β_4). The deletion of this path from Model A resulted in a large increase in chi-square (Table 8). Thus, the hypothesis that cognitive deficits in the elderly can be attributed primarily to lack of physical exercise was not supported.

We next investigated the role of health. Researchers have suggested that health may play a more important role in cognitive deficits than chronological age (Steuer & Jarvik, 1981). We first added to Model A a direct path from health to each performance variable (β_6). The resulting model, Model C, did not fit the data significantly better than did Model A for any performance variable, and the added parameter (β_6) did not reach significance. Also, the addition of β_6 did not cause the paths from age to performance or from exercise to performance to decrease in magnitude. These results suggest that there was no direct effect of health on performance. Next, we deleted the

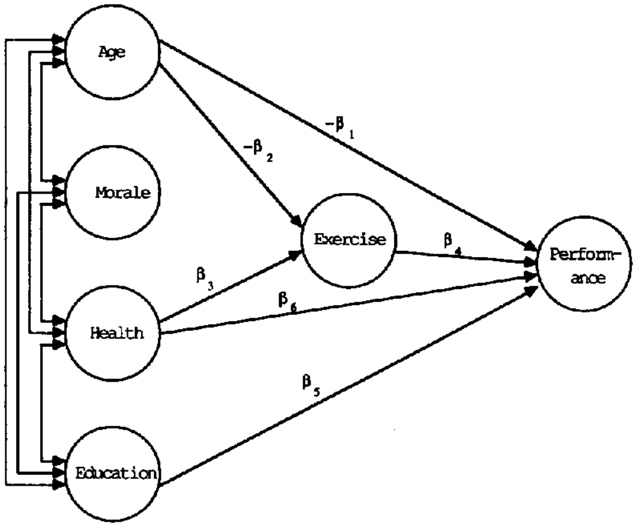

Figure 2. Alternate pathways for competing models of reasoning and working memory. (For model of reaction time, β_1 was positive, and β_4 was negative. Paths β_1 through β_4 were included in all models. β_5 was included only in the model for reasoning. β_6 was not required in Model A but was included in several model comparisons. Nonsignificant paths from education and morale to exercise were omitted from the diagram. Single-headed arrows indicate hypothesized unidirectional influences; double-headed arrows indicate correlations with no specified direction of influence.)

Table 7
Statistical Values for Alternate Models (N = 300)

Model	χ^2	*df*	*p*	ρ
		Reaction time		
Msat	62.42	62	.461	1.000
A	67.65	65	.387	0.997
B	99.23	66	.005	0.967
C	63.94	64	.487	1.000
D	95.65	65	.008	0.969
E	79.44	66	.124	0.986
F	70.12	65	.310	0.995
		Working memory		
Msat	71.98	62	.181	0.991
A	77.60	65	.136	0.989
B	99.23	66	.005	0.970
C	76.65	64	.134	0.988
D	97.78	65	.005	0.970
E	84.62	66	.061	0.983
F	81.20	65	.085	0.985
		Reasoning		
Msat	65.15	60	.302	0.996
A	65.58	62	.354	0.997
B	93.93	63	.007	0.971
C	65.25	61	.331	0.978
D	93.63	62	.006	0.997
E	77.55	63	.103	0.990
F	74.53	62	.132	0.991

Note. Msat = fully saturated model; A = best fitting model; B = causal pathway β_1 deleted from Model A; C = causal pathway β_6 added to Model A; D = causal pathway β_1 deleted from Model C; E = causal pathway β_4 deleted from Model A; F = causal pathway β_4 deleted from Model C.

path from age to performance (β_1) from Model C, resulting in Model D, in which health influenced performance both directly and indirectly through physical exercise and in which there was no direct effect of age. This last deletion resulted in a significant increase in chi-square, suggesting that age-related deficits in cognition cannot be completely explained in terms of impaired health and decreased physical activity.

Finally, we investigated the central hypothesis of our study, that exercise has a positive effect on cognitive performance. To test this hypothesis, we deleted the direct path from exercise to each performance variable (β_4) from Model A. For the resulting model, Model E, there was a significant decrease in model fit. Because this deletion also eliminated the indirect effect of health on performance mediated through exercise, we next deleted the path from exercise to performance (β_4) from Model C, to which a direct path (β_6) from health to performance had been added. This left Model F, in which health and age affected both exercise and the performance variables directly, but in which there was no effect of exercise on performance. As shown in Table 8, the deletion of the paths from exercise to the performance variables (β_4) in Model C resulted in significant increases in chi-square, suggesting an effect of exercise on performance that is independent of health.

Discussion

Models showing relationships of age, health, education, morale, and exercise with RT, working memory, and reasoning proved to fit the data well. Performance variables in each model were directly influenced by age and by exercise. For the model of reasoning, there was an additional causal pathway, from education to reasoning. In each of the three models, the amount and vigor of exercise in which individuals engaged was a negative function of age and a positive function of health. Contrary to some previous research (Ostrow, 1984; Stones & Kozma, 1985), no relationship of morale or education with exercise was seen. Our results concur, however, with those of Milligan et al. (1984) in finding no effect of morale on performance variables and with those of previous researchers in finding that morale correlated with both health and education. Despite a slight drop in significance in the path from exercise to working memory when gender was included as an exogenous variable, we believe that the same models are appropriate for men and women.

The main hypothesis of a positive relationship between physical exercise and cognition in older people was supported. The fit of the models was significantly decreased when pathways leading from exercise to the performance variables (β_4) were deleted. However, the results did not support the stronger version of the hypothesis proposed by Bortz (1982), that age-re-

Table 8
Comparison of Competing Models With Best Fitting Model (N = 300)

	Response speed		Working memory		Reasoning	
Comparison	$\Delta\chi^2$	*df*	$\Delta\chi^2$	*df*	$\Delta\chi^2$	*df*
Model A – Msat	5.23	3	5.62	3	0.43	2
Model B – Model A	31.58***	1	21.63***	1	28.35***	1
Model A – Model C	3.71	1	.95	1	.33	1
Model D – Model C	31.71***	1	21.13***	1	28.38***	1
Model E – Model A	11.79***	1	7.02**	1	11.97***	1
Model F – Model C	6.18*	1	4.55*	1	9.28**	1

Note. Msat = fully saturated model; Model A = best fitting model; Model B = causal pathway β_1 deleted from Model A; Model C = causal pathway β_6 added to Model A; Model D = causal pathway β_1 deleted from Model C; Model E = causal pathway β_4 deleted from Model A; Model F = causal pathway β_4 deleted from Model C.
* $p < .05$. ** $p < .01$. *** $p < .001$.

lated cognitive deficits were primarily the result of the physically inactive life that many older people lead. Large decreases in model fit occurred when the path from age to each performance variable (β_1) was deleted.

The notion that age-related deficits in cognition could be attributed largely to deteriorating health was not supported. The addition of a direct pathway from health to performance variables (β_6) did not improve the fit of the model. Deleting the exercise-to-performance path (β_4) from the model after adding the health-to-performance path (β_6) caused a significant decrease in fit of the models for each of the performance variables. Even with the addition of the health-to-performance path (β_6) to Model A, the paths from exercise to performance (β_4) remained significant for each performance variable. These results suggest that there was an effect of exercise that was independent of the effect of health and that, for this sample at least, the only effect of health on performance was mediated through its effect on exercise. However, the sample for this study was unusually healthy, and the relationships we found might not generalize to a sample more representative of the age group.

Because the most consistent physiological effect of exercise is an increase in cardiovascular-respiratory fitness, it has been assumed that any benefit of exercise on cognition is mediated through fitness. Our original intention was to include fitness as an intervening variable between exercise and performance. Maximum oxygen uptake ($\dot{V}_{O_{2max}}$) is generally considered the best measure of fitness and one that is highly correlated with physical exercise. However, testing participants for $\dot{V}_{O_{2max}}$ proved impractical in this study. The measures we used—heart rate, blood pressure, and vital capacity—were not satisfactory indicators of a latent fitness variable. One reason may be that several participants were taking medication that regulated heart rate and blood pressure. Also, it is doubtful that blood pressure is linearly related to either cognition or health, in that very low blood pressure is no more optimal than high blood pressure. A positive relationship between moderately elevated blood pressure and cognitive performance has been reported (Hertzog et al., 1978; Wilkie & Eisdorfer, 1971). Of our measures, vital capacity was the most reliable measure of fitness. However, rather than rely on one imperfect measure of respiratory fitness, we excluded fitness from the model.

In this study, we have added to the findings that are consistent with a causal relationship between exercise and cognition, but we could not unequivocally establish such a relationship. Even though the models tested in the study suggest that exercise has a beneficial effect on cognition, it cannot be denied that there may be other models that would fit the data equally well. Although studies of chronic exercisers may lack the precision of experimental studies, we believe that because of their ecological validity, studies of chronic exercisers have an important part to play in the investigation of the relationship between exercise and cognitive performance. A logical next step that could provide stronger support for a causal relationship would be a longitudinal study.

References

Abourezk, T. (1988). The effects of regular aerobic exercise on short-term memory efficiency in the older adult. In A. C. Ostrow (Ed.), *Aging and motor behavior* (pp. 105–113). Indianapolis, IN: Benchmark.

Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin, 103,* 411–423.

Barnhart, E. R. (Ed.). (1986). *Physicians' desk reference* (40th ed.). Oradell, NJ: Medical Economics Company.

Baylor, A. M., & Spirduso, W. W. (1988). Systematic aerobic exercise and components of reaction time in older women. *Journal of Gerontology: Psychological Sciences, 43,* P121–P126.

Bentler, P. M. (1980). Multivariate analysis with latent variables: Causal modeling. *Annual Review of Psychology, 31,* 419–456.

Bentler, P. M. (1988). Causal modeling via structural equation systems. In J. R. Nesselroade & R. B. Cattell (Eds.), *Handbook of multivariate experimental psychology* (2nd ed., pp. 317–335). New York: Plenum Press.

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88,* 588–606.

Blumenthal, J. A., & Madden, D. J. (1988). Effects of aerobic exercise training, age, and physical fitness on memory-search performance. *Psychology and Aging, 3,* 280–285.

Bollen, K. A. (1986). Sample size and Bentler and Bonett's nonnormed fit index. *Psychometrika, 51,* 375–377.

Bortz, W. M. (1982). Disuse and aging. *Journal of the American Medical Association, 248,* 1203–1208.

Bradburn, N. M., & Caplovitz, D. (1965). *Reports on happiness.* Chicago: Aldine.

Charness, N. (1985). Aging and problem-solving performance. In N. Charness (Ed.), *Aging and human performance* (pp. 225–259). New York: Wiley.

Clarkson-Smith, L. (1985). The components of verbal analogies and their relations to fluid and crystallized intelligence (Doctoral dissertation, University of California, Riverside, 1984). *Dissertation Abstracts International, 45,* 2329B.

Clarkson-Smith, L., & Hartley, A. A. (1989). Relationships between physical exercise and cognitive abilities in older adults. *Psychology and Aging, 4,* 183–189.

Craik, F. I. M., & Byrd, M. (1982). Aging and cognitive deficits: The role of attentional resources. In F. I. M. Craik & S. Trehub (Eds.), *Aging and cognitive processes* (pp. 191–211). New York: Plenum Press.

Crawford, J., & Stankov, L. (1983). Fluid and crystallized intelligence and primacy/recency components of short-term memory. *Intelligence, 7,* 227–252.

Cunningham, D. A., Rechnitzer, P. A., Howard, J. H., & Donner, A. P. (1987). Exercise training of men at retirement: A clinical trial. *Journal of Gerontology, 42,* B17–B23.

Daneman, M., & Carpenter, P. A. (1980). Individual differences in working memory and reading. *Journal of Verbal Learning and Verbal Behavior, 19,* 450–466.

deVries, H. A. (1983). Physiology of exercise and aging. In D. S. Woodruff & J. E. Birren (Eds.), *Aging: Scientific perspectives and social issues* (2nd ed., pp. 285–304). Monterey, CA: Brooks/Cole.

Dillon, W. R., Kumar, A., & Mulani, N. (1987). Offending estimates in covariance structure analysis: Comments on the causes and solutions to Heywood cases. *Psychological Bulletin, 101,* 126–135.

Doan, R. E., & Scherman, A. (1987). The therapeutic effect of physical fitness on measures of personality: A literature review. *Journal of Counseling and Development, 66,* 28–36.

Dustman, R. E., Ruhling, R. O., Russell, E. M., Shearer, D. E., Bonekat, H. W., Shigeoka, J. W., Wood, J. W., & Bradford, D. C. (1984). Aerobic exercise training and improved neuropsychological function of older adults. *Neurobiology of Aging, 5,* 35–42.

Elsayad, M., Ismail, A. H., & Young, R. J. (1980). Intellectual differences of adult men related to age and physical fitness before and after an exercise program. *Journal of Gerontology, 35,* P383–P387.

Hasher, L., & Zacks, R. T. (1979). Automatic and effortful processes in memory. *Journal of Experimental Psychology: General, 108,* 356–388.

Hayduk, L. A. (1987). *Structural equation modeling with LISREL: Essentials and advances.* Baltimore, MD: Johns Hopkins University Press.

Hertzog, C., Schaie, K. W., & Gribbin, K. (1978). Cardiovascular disease and changes in intellectual functioning from middle to old age. *Journal of Gerontology, 33,* 872–883.

Heyward, V. H. (1984). *Designs for fitness: A guide to physical fitness appraisal and exercise prescription.* Minneapolis, MN: Burgess.

Horn, J. (1975). *Gf-Gc sampler.* Unpublished manuscript, University of Denver, Psychology Department.

Horn, J. (1980). Concepts of intellect in relations to learning and adult development. *Intelligence, 4,* 285–317.

Jöreskog, K. G., & Sörbom, D. (1984). *LISREL VI: Analysis of linear structural relationships by the method of maximum likelihood.* Mooresville, IN: Scientific Software.

Madden, D. J., Blumenthal, J. A., Allen, P. A., & Emery, C. F. (1989). Improving aerobic capacity in healthy older adults does not necessarily lead to improved cognitive performance. *Psychology and Aging, 4,* 307–320.

Manton, K. G., Siegler, I. C., & Woodbury, M. A. (1986). Patterns of intellectual development in later life. *Journal of Gerontology, 41,* P486–P499.

Marsh, H. W., Balla, J. R., & McDonald, R. P. (1988). Goodness-of-fit indices in confirmatory factor analysis: The effect of sample size. *Psychological Bulletin, 103,* 391–410.

Milligan, W. L., Powell, D. A., Harley, C., & Furchtgott, E. (1984). A comparison of physical health and psychosocial variables as predictors of reaction time and serial learning performance in men. *Journal of Gerontology, 39,* 704–710.

Neugarten, B. I., Havighurst, R. J., & Tobin, S. S. (1961). The measurement of life satisfaction. *Journal of Gerontology, 16,* 134–143.

Ohlsson, M. (1975). An experimental study on physical fitness related to information processing in elderly people. In G. Borg (Ed.), *Physical work and effort: Proceedings of the first international symposium, Wenner-Gren Center, Stockholm* (pp. 137–143). New York: Pergamon Press.

Ostrow, A. C. (1984). *Physical activity and the older adult: Psychological perspectives.* Princeton, NJ: Princeton Book Company.

Powell, R. R., & Pohndorf, R. H. (1971). Comparison of adult exercisers and non-exercisers on fluid intelligence and selected physiological variables. *Research Quarterly, 42,* 70–77.

Raven, J. C. (1972). *Advanced progressive matrices: Set I.* London: Silver End Press.

Raven, J. C. (1974). *Advanced progressive matrices: Set II.* London: Silver End Press.

Reese, H. W., & Rodeheaver, D. (1985). Problem solving and complex decision making. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 474–499). New York: Van Nostrand Reinhold.

Salthouse, T. A. (1985). Speed of behavior and its implications for cognition. In J. E. Birren & K. W. Schaie (Eds.), *Handbook of the psychology of aging* (2nd ed., pp. 400–416). New York: Van Nostrand Reinhold.

Schultz, N. R., Elias, M. F., Robbins, M. A., Streeten, D. H., & Blakeman, N. (1986). A longitudinal comparison of hypertensives and normotensives on the Wechsler Adult Intelligence Scale: Initial findings. *Journal of Gerontology, 41,* P169–P175.

Spirduso, W. L. (1980). Physical fitness, aging, and psychomotor speed: A review. *Journal of Gerontology, 35,* P850–P865.

Steuer, J., & Jarvik, L. F. (1981). Cognitive functioning in the elderly: Influence of physical health. In J. L. McGaugh & S. B. Kiesler (Eds.), *Aging: Biology and behavior* (pp. 231–253). New York: Academic Press.

Stones, M. J., & Kozma, A. (1985). Physical performance. In N. Charness (Ed.), *Aging and human performance* (pp. 261–291). New York: Wiley.

Wechsler, D. (1955). *Wechsler Adult Intelligence Scale.* New York: The Psychological Corporation.

Wilkie, F. L., & Eisdorfer, C. (1971). Intelligence and blood pressure in the aged. *Science, 172,* 959–962.

Wood, V., Wylie, M., & Sheafer, B. (1969). An analysis of a short self-report measure of life satisfaction: Correlation with rater judgements. *Journal of Gerontology, 24,* P465–P469.

Received June 19, 1989
Revision received March 12, 1990
Accepted March 16, 1990 ■