DRS PRICE MF-$0.83 HC-$1.67 Plus Postage.
ESCRIPTORS *College Entrance Examinations; *Factor Analysis; Higher Education; *Item Analysis; *Law Schools; *Predictive Validity; Professional Education; Test Validity
DENTIFIERS *Law School Admission Test

BSTRACT

A factor analysis of the Law School Admission Test LSAT) battery was undertaken to gain a better understanding of the pecific abilities which contribute to performance on the tests. To etermine whether greater amounts of testing time should be allocated o groups of items defined by the factor analysis, a validity study as conducted. Existing section of the LSAT, the morning test, were hown to correspond closely to the factors. Variance in the afternoon ests was adequately explained by LSAT factors. It was shown that the ength of the test battery could be reduced to a half day of testing ime without impairing predictive validity. (Author/BW)

Factor Analysis and Validity Study of the Law School Admission Test Battery[1]

Alfred B. Carlson
Educational Testing Service

A factor analysis of the Law School Admission Test was authorized by the Law School Admission Council in 1966. The purpose was to gain a better understanding of the specific abilities which contribute to performance on the test.

When this study was conducted the LSAT battery comprised a full day of testing. The morning session was devoted to the LSAT, which was composed of five separately timed sections--Reading Comprehension, Data Interpretation, Reading Recall, Principles and Cases, and Figure Classification. The afternoon session included Writing Ability, composed of three separately timed sections--Error Recognition, Organization of Ideas, and Editing--and General Background. These tests contained 150, 100, and 90 items respectively. The two research questions to be answered were: (1) What are the abilities measured by the morning test? (Factor Analysis). (2) Are the abilities measured by the afternoon tests different from those measured by the morning test? (Factor Extension). In addition, it was hoped that a classification of items based on abilities would provide a useful supplement or alternative to the present classification based on item type and would provide guidance for the writing and/or analysis of future items and for future studies of the structure of the test.

To determine whether greater amounts of testing time should be allocated to groups of items defined by the factor analysis study, a follow-up validity study was also conducted. During the conduct of the study the Council became particularly interested in restructuring and shortening the LSAT battery; thus, many of the analyses were directed toward achieving these specific alternatives.

---

[1] This paper represents the final report of a study begun in 1966 and sponsored by the Law School Admission Council. Earlier brief reports of the research were issued in the LSAC Annual Reports of 1967, 1968, and 1970. This final report appears as LSAC-70-3 in Law School Admission Research, Volume II, 1970 - 1974 and served as the basis for a paper entitled "Analyzing and Restructuring a National Admissions Test" given at the annual meeting of the National Conference on Measurement in Education, Washington, D. C., 1975.

The research was carried out in three parts: (1) a factor analysis of the LSAT; (2) an extension of the LSAT factors into the Writing Ability and General Background space; (3) a validity study and optimum timing allocation of part scores.

The sample for the factor analysis and factor extension was 13,676 students who had taken the Law School Admission Test battery in November 1965. The sample for the validity study consisted of 941 of these students who had been admitted to one of 18 law schools.

The Factor Analysis

Factor analysis is a systematic way of defining groups of items toward which candidates tend to behave in the same way and which are probably measuring the same ability. As such, it is concerned with the relationships between measures—in this case, items. If several items are interrelated, they may all, at least in part, be measuring some common ability. If so, then this ability "explains" the relationship between the items. Factor analysis is a statistical procedure that conceptually, at least, groups these items measuring a common ability and then defines a hypothetical measure of this ability giving greatest weight to the items which measure this ability "best" and less weight to those items which do not measure it as well. Statistically, common measures, or factors, can then be "removed" from the items, and the relationships then remaining between the items can be examined. Factor analysis may then be thought of as a procedure for minimizing the remaining (or residual) relationships by means of as small a number of factors as possible.

Factor analysis as applied to this study may be considered as either a hypothesis-confirming technique (i.e., the current sectioning of the test would be confirmed); or a hypothesis-generating technique (i.e., a different organization of the items in the test would be suggested).

The results of a factor analysis are factor loadings. These loadings may be interpreted as the correlations between the items and the factor; thus, the factor loading for an item on a factor may be considered the factorial validity of that item. The factor loadings are then used to interpret the factor since interpretation is subjective.

Each of the 150 test items was considered a variable in this study. Since the test is scored for the number of "rights only", correct answers are scored 1, and incorrect answers, omitted items, and items not reached are scored 0.

The complete matrix of interitem tetrachoric correlation coefficients was computed. The communality of each item was then estimated using a modification of the highest correlation procedure suggested by Tucker (Carlson, 1967). Using these estimates of the communalities, principal axis solutions were found for the 75 odd-numbered items and for the 75 even-numbered items. The number of factors to rotate in each case was determined by examining the difference between successive characteristic roots. The break between the roots in both instances was between the ninth and tenth.

Each set was then rotated to a Varimax solution. The rotated factors from the two analyses were subjectively matched. The loadings for the matched factors were used as estimates of the factor loadings for the 150 items to be iterated by a procedure suggested by Boldt (1965) that uses the method of steepest dissent to minimize the off-diagonal residual correlation matrix. Several different numbers of factors were iterated.

The final decision as to the appropriate number of factors was based upon all available information, including the interpretability of the factors. The seven factoral solution was chosen and rotated to the Varimax solution.

Six of the seven factors were tentatively interpreted as I Reading Comprehension (RC), II Verbal Inductive Reasoning (VIR), III Tabular Data Interpretation (TDI), IV Figure Classification (FC), V Graphical Data Interpretation (GDI) and VII Numerical Recall (NR). The seventh factor, VI (P&C), involved one case from the Principles and Cases section and may be peculiar to this form of the test. The rotated factor loadings for each test item are shown in Table 1.

Table 1

Rotated Factor Loadings

Item Type	Item Number	Factor RC I	VIR II	TDI III	FC IV	GDI V	P&C VI	NR VII
Reading Comprehension								
First Passage	1	.42						
	2	.57						
	3	.48						
	4	.38						
	5	.48						
	6	.45						
Second Passage	7	.31						
	8	.44						
	9	.31						
	10	.40						
	11	.50						
	12	.43						
Third Passage	13	.48						
	14	.44						
	15	.26						
	16	.47						
	17							
	18	.40						
	19	.34						
Fourth Passage	20	.44	.26					
	21	.48						
	22	.43	.21					
	23	.34	.22					
	24	.54	.33			.22		
	25	.32	.24					

Table 1 Continued

Rotated Factor Loadings

Item Type	Item Number	Factor RC I	VIR II	TDI III	FC IV	GDI V	P&C VI	NR VII
Data Interpretation								
First Data Set	26	.25		.34				
	27			.22				
	28			.42				
	29	.29		.39				
	30			.33				
	31			.35				
	32			.45				
Second Data Set	33							
	34			.39				
	35							
	36			.28				
	37			.23				
	38			.20				
	39							
Third Data Set	40			.34				
	41			.51				
	42			.35				
	43			.47				
	44			.52				
	45			.37				
	46			.52		.21		
Fourth Data Set	47			.25				
	48							
	49			.25		.38		
	50	.21		.30		.39		
	51	.22		.30		.48		
	52			.32		.28		
	53			.27		.42		
Fifth Data Set	54					.53		
	55					.39		
	56			.20		.42		
	57		.22			.67		
	58		.23			.63		
	59					.59		
	60					.53		

Table 1 Continued

Rotated Factor Loadings

Item Type	Item Number	Factor RC I	VIR II	TDI III	FC IV	GDI V	P&C VI	NR VII
Reading Recall								
Passage 1	61	.35		.21				.20
	62	.30						
	63	.21						
	64							
	65	.26						
	66							
	67	.27						
	68	.21						
	69							
	70	.23						
Passage 2	71	.44						
	72	.35						
	73	.40						
	74	.48						
	75							
	76	.29						
	77	.45						
	78	.45		.20				
	79	.3						
	80	.38						
Passage 3	81	.36		.23				
	82	.21		.22				.48
	83			.21				.50
	84	.23						
	85							
	86	.35						
	87	.26						.25
	88							.51
	89	.25	.28					.43
	90							.24

Table 1 Continued

Rotated Factor Loadings

Item Type	Item Number	Factor RC I	VIR II	TDI III	FC IV	GDI V	P&C VI	NR VII
Principles and Cases								
Part A, Case 1	91							
	92							
	93	.25						
	94						.22	
	95	.23						
Case 2	96							
	97	.21						
	98							
	99							
Case 3	100						.60	
	101						.64	
	102						.62	
	103						-.60	
Part B, Principle 1	104							
	105							
	106							
	107							
Principle 2	108	.30		.23				
	109							
	110	.28						
	111							
	112	.23						
Principle 3	113							
	114							
	115							
	116							
Part C, Group 1	117							
	118		.32					
	119		.49					
	120		.52					
Group 2	121		.65					
	122		.41					
	123		.57					
	124		.66					
	125		.67					
Group 3	126		.55					
	127		.62					
	128		.64					
	129		.49					
	130		.50					

Table 1 Continued

Rotated Factor Loadings

Item Type	Item Number	Factor RC I	VIR II	TDI III	FC IV	GDI V	P&C VI	NR VII
Figure Classification								
	131				.53			
	132				.57			
	133	.20			.75			
	134				.69			
	135	.21			.51			
	136				.43			
	137				.33			
	138				.29			
	139				.46			
	140				.36			
	141							
	142				.38			
	143				.36			
	144				.28			
	145				.24			
	146	.20		.23	.35			
	147							
	148				.21			
	149				.22			
	150				.22			

and the relationships remaining between the items were studied. Of the seven factors identified in the morning LSAT, only the first, Reading Comprehension, was measured by the Writing Ability Test; and only the first and the second, Verbal Inductive Reasoning, were measured by the General Background Test. Table 2 shows the estimated factor loadings for Writing Ability. The estimated factor loadings for General Background are shown in Table 3. There were no residual correlations of 0.20 or greater between sections or tests. The removal of the factors from the morning test lowered the root mean square correlation from 0.10 to 0.04. It was decided not to factor the afternoon tests further. (With an average correlation between items of 0.04, it is doubtful that any large factors remain.) The conclusions were that the Writing Ability and General Background tests measure primarily verbal ability, as indicated by their loadings on the reading comprehension factor; that some items of the General Background Test also measure Verbal Inductive Reasoning; and that the afternoon tests probably do not provide significant measurement of abilities other than those measured by the morning test.

The Validity Study

The test scores analyzed in the validity study were (1) section scores based upon existing separately timed sections of the test), (2) factorially defined scores (based upon items found in the factor analysis to be related to single factor), and (3) selected scores (clusters of items from the Writing bility test remaining after the morning test factors had been removed and two ubsections of Principles and Cases which were not related to any of the factors). n addition to test data, both undergraduate average (UGA) and first-year average FYA) grades in law school were available for these students. So that grades ould be pooled across law schools, both FYA and UGA were standardized by etting the within-law school mean and standard deviation for each, equal to) and 10 respectively. FYA was also scaled using a weighted composite of SAT and WA sections as the anchor. The validity of each of the section,

Table 2

Estimated Factor Loadings for Writing Ability*

Item Type	Item No.	Factor** I
Error Recognition		
	1	.36
	2	.27
	3	.28
	4	.23
	5	.27
	6	
	7	.22
	8	.21
	9	
	10	.33
	11	.28
	12	.34
	13	.21
	14	.28
	15	
	16	.26
	17	.32
	18	.37
	19	.37
	20	.31
	21	.21
	22	.29
	23	.38
	24	.21
	25	
	26	.22
	27	
	28	.27
	29	
	30	
	31	.37
	32	
	33	
	34	.25
	35	.30

Item Type	Item No.	Factor** I
Organization of Ideas		
Set A		
	36	.25
	37	
	38	
	39	.29
	40	
	41	
	42	.27
	43	
	44	
	45	
Set B		
	46	.44
	47	
	48	
	49	.28
	50	.31
	51	
	52	.23
	53	
	54	.20
	55	.29
Set C		
	56	
	57	.23
	58	
	59	.28
	60	
	61	.26
	62	
	63	
	64	

Item Type	Item No.	Factor** I
Editing		
	65	
	66	.20
	67	.26
	68	
	69	
	70	
	71	
	72	.2[illegible]
	73	
	74	.20
	75	
	76	.21
	77	.31
	78	.34
	79	
	80	.29
	81	
	82	
	83	
	84	.25
	85	
	86	
	87	
	88	.26
	89	.29
	90	.23
	91	
	92	
	93	
	94	.21
	95	
	96	.28
	97	.23
	98	.23
	99	
	100	

*Factor loadings of less than 0.20 are omitted.

**There are no loadings of 0.20 or greater on factors II through VII.

Table 3

Estimated Factor Loadings for General Background*

Item Number	Factor** I	Factor** II	Item Number	Factor** I	Factor** II	Item Number	Factor** I	Factor** II
1			31	.34	.24	61	.25	.22
2			32	.21	.20	62	.24	
3	.29		33	.32		63	.28	
4	.30		34	.20		64	.26	
5	.27		35			65	.20	
6	.49		36			66	.32	
7	.28		37	.45		67	.24	
8	.44		38			68	.33	
9	.32		39			69	.27	
10	.28		40	.25		70	.26	
11			41	.28		71		
12	.33		42	.20		72	.21	
13			43	.27		73		
14	.27		44	.32		74	.25	
15	.36		45	.29		75	.24	
16			46	.24		76		
17	.25		47			77	.22	
18	.29		48	.31		78	.43	
19	.25		49		.28	79	.32	.22
20			50	.33		80	.26	.20
21	.26		51	.21		81	.35	
22	.36	.30	52	.29		82		
23	.26		53	.25		83	.26	
24	.36		54			84	.26	
25	.27		55	.31	.33	85		
26	.24		56	.34		86		
27	.41		57			87	.27	
28	.24		58	.23		88		
29			59			89	.29	
30	.31		60	.35		90	.21	

*Factor loadings of less than 0.20 are omitted.

**There are no loadings of 0.20 or greater on factors III through VII.

factor, and selected scores used in this study; the number of items contributing to each score, and the correlations between scores within each group are given in Table 4.

Optimal timing allocations were computed using a procedure, developed by Jackson and Novick (1969), which assigns lengths to the tests of a battery so as to maximize the correlation with a given criterion when the total testing time is fixed. The criterion used for these analyses was scaled FYA. For many of these analyses, the total sample was randomly divided into two samples. Time allocations were computed for each sample, and the results applied to the other sample to compute the validity coefficients. The result of this kind of analysis is the specification of the amount of testing time that should be given to each section or item type in order to maximize the predictive validity of the total test. Table 5 gives the optimal testing times and validities for the five sections of the morning test when total testing time is 175 minutes. Table 6 gives the optimal testing times and validities for the factorially defined scores and selected scores. The analyses leading to the results reported in Tables 7 and 8 are addressed to the possibility of shortening the test to a half day of testing.

Correlation and regression analyses showed Reading Comprehension, both the factorially defined score and the section, to be the most valid of the item types. Figure Classification was the least valid factorially defined score, and General Background was the least valid section. Graphical Data Interpretation had somewhat greater predictive validity than Tabular Data Interpretation. The selected score, Error Recognition items characterized by problems in diction, had greater validity than those with problems in verbosity. The most valid pair of factorially defined scores was Reading Comprehension and Graphical Data Interpretation. The most valid pair of sections was Reading Comprehension and Data Interpretation. General Background was found to make no useful contribution to the predictive validity of the battery, and is not included in any of the time allocation analyses.

The operational sections of the morning test required 175 minutes of testing time. This was broken down into: Reading Comprehension, 30 minutes; Data Interpretation, 45 minutes; Reading Recall, 30 minutes; Principles and

Table 4

LSAT Section and Factor Score Correlations

A.M. and P.M. Sections:

Score	#Items	Std. r_{FYA}	Scaled r_{FYA}	Intercorrelations								
				R.C.	D.I.	R.R.	P.C.	F.C.	E.R.	O.I.	E.	G.B.
Reading Comprehension	25	.19	.34		.36	.53	.43	.18	.49	.32	.41	.55
Data Interpretation	35	.16	.29			.46	.37	.39	.34	.30	.24	.37
Reading Recall	30	.16	.30				.35	.20	.45	.30	.40	.46
Principles and Cases	40	.19	.30					.17	.29	.26	.25	.32
Figure Classification	20	.08	.14						.26	.18	.17	.21
WA, Error Recognition	35	.13	.26							.36	.45	.53
WA, Organization of Ideas	30	.12	.21								.29	.32
WA, Editing	35	.11	.23									.40
General Background	90	.08	.23									

Factors and selected scores:

Score	#Items	Std. r_{FYA}	Scaled r_{FYA}	Intercorrelations									
				R.C.	V.I.R.	T.D.I.	F.C.	G.D.I.	Dic.	Ver.	A	B	C
Reading Comprehension	22	.17	.32		.21	.28	.22	.21	.36	.36	.27	.34	.22
Verbal Inductive Reasoning	13	.11	.16			.16	.02	.30	.08	.12	-.01	.14	.99
Tabular Data Interpretation	12	.10	.20				.32	.30	.16	.18	.14	.19	.16
Figure Classification	11	.06	.13					.20	.15	.19	.08	.13	.02
Graphical Data Interpretation	8	.13	.21						.16	.18	.07	.18	.29
Diction	9	.15	.24							.29	.21	.19	.08
Verbosity	9	.09	.18								.14	.13	.12
Principles + Cases, Part A	13	.14	.21									.18	-.00
Principles + Cases, Part B	13	.13	.22										.15
Principles + Cases, Part C	14	.11	.16										

Table 5

Optimal Testing Times and Validity for Morning Test Sections

(Each Total Time = 175 minutes)

		Altered Testing Times					
Variables		Sample 1		Sample 2		Total	
Reading Comprehension		80	80	60	60	72	55
Data Interpretation		47	45	20	20	38	40
Reading Recall		30	30	—	—	7	20
Principles and Cases		18	20	84	85	58	60
Figure Classification		--	—	11	10	--	--
Correlation with Scaled FYA	Computation Sample	.42	.42	.42	.42	.42	.41
	Validation Sample	.40	.40	.39	.39	--	--

Table 6

Optimal Testing Times and Validity for Factors and Selected Scores

(Each Total Time = 175 minutes)

		Altered Testing Times					
Variables		Sample 1		Sample 2		Total	
Reading Comprehension		49	50	—	—	27	30
Verbal Inductive Reasoning		7	10	—	—	11	10
Tabular Data Interpretation		—	--	11	10	7	10
Figure Classification		--	—	—	—	—	—
Graphical Data Interpretation		25	25	4	5	14	10
Principles and Cases, Part A		55	50	--	--	39	40
Principles and Cases, Part B		--	—	134	135	52	50
Diction		29	30	26	25	25	25
Verbosity		10	10	--	--	--	--
Correlation with Scaled FYA	Computation Sample	.43	.43	.46	.46	.42	.42
	Validation Sample	.39	.39	.36	.36	—	—

Table 7

Optimal Testing Times and Validity for
Morning Test and Writing Ability Sections

(Each Total Time = 175 minutes)

		Altered Testing Times					
Variables		Sample 1		Sample 2		Total	
Reading Comprehension		58	55	59	60	67	65
Data Interpretation		43	40	19	20	35	30
Reading Recall		18	20	--	--	5	--
Principles and Cases		14	20	83	30	57	55
Figure Classification		5	--	11	15	--	--
WA, Error Recognition		27	25	--	--	5	10
WA, Organization of Ideas		6	15	3	--	6	15
WA, Editing		4	--	--	--	--	--
Correlation with Scaled FYA	Computation Sample	.42	.42	.42	.42	.41	.41
	Validation Sample	.37	.38	.39	.39	--	--

Table 8

Optimal Testing Times and Validity for
Morning Test Sections I-IV and Writing Ability Sections I & III

(Each Total Time = 175 minutes)

		Altered Testing Times					
Variables		Sample 1		Sample 2		Total	
Reading Comprehension		61	55	62	60	67	50
Data Interpretation		44	40	30	30	36	35
Reading Recall		18	20	--	--	5	20
Principles and Cases		16	20	83	85	59	60
WA, Error Recognition		28	25	--	--	7	10
WA, Editing		8	15	--	--	--	--
Correlation with Scaled FYA	Computation Sample	.42	.42	.42	.42	.41	.41
	Validation Sample	.37	.37	.40	.40	--	--

Cases, 55 minutes; and Figure Classification, 15 minutes. There were rather noticeable differences in the solutions obtained for each sample. For this reason analyses based on the total sample were also computed. The total sample gave a better estimate of the most valid time allocation for the population, but the validity of a given allocation could not be estimated since the total sample had now been used to estimate the timings.

It was concluded that:

1) the validity of factorially defined scores and selected scores was not appreciably greater than that of scores from existing sections of the test;

2) reallocating testing time among factorially defined scores and selected scores resulted in a negligible gain over current predictive validity;

3) approximately optimum lengths were assigned to existing sections in the morning test; and

4) General Background made no useful contribution to the predictive validity of the battery.

It was suggested that testing time for the LSAT battery could be reduced to a morning without a significant decrease in predictive validity by (1) omitting General Background, (2) omitting one of the three Writing Ability sections, Organization of Ideas, and reducing the time of another section, Editing, from 30 to 20 minutes and (3) omitting the Figure Classification section of the morning test. Table 9 shows the resulting validities for the total sample when several combinations of arbitrary testing times were used for the Morning Test and Writing Ability sections.

The recommended changes in the test battery were approved by the Law School Admission Council. Consequently, students devote only a half day to taking the test, and testing fees have been lowered.

Table 9

Total Sample Validity for Arbitrary Testing Times
Using Morning Test and Writing Ability Sections

(Each Total Time = 200 minutes)

Variables	Altered Testing Times					
Reading Comprehension	30	30	60	60	70	70
Data Interpretation	45	45	45	45	35	35
Reading Recall	30	30	--	--	--	--
Principles and Cases	55	55	55	55	55	55
Figure Classification	--	--	--	--	--	--
WA, Error Recognition	20	20	20	20	20	20
WA, Organization of Ideas	--	20	--	20	--	20
WA, Editing	20	--	20	--	20	--
Correlations with Scaled FYA	.40	.40	.41	.41	.41	.41

composite as a function of relative component lengths with a fixed total testing time (ETS RB-69-14). Princeton, N.J.: Educational Testing Service, February 1969.

Alfred B. Carlson
1970