דוח המרכז

324

How and How Well Do Schools of Architecture Select Their Students:
Israel's Technion as a Case Study

Carmel Oren
Ayelet Cohen
Gabriela Goldschmidt
Rachel Sebba

August 2005

מרכז ארצי לבחינות ולהערכה (ע"ר)
NATIONAL INSTITUTE FOR TESTING & EVALUATION
מיסודן של האוניברסיטאות בישראל
ת"ד 26015 ירושלים 91260, טל' 02-6759555

דוח מרכז 324

ISBN:965-502-114-9

How and How Well Do Schools of Architecture Select Their Students: Israel's Technion as a Case Study

NCME Annual Meeting, Chicago, April 2003

Carmel Oren* (corresponding author: carmel@nite.org.il)

Ayelet Cohen*

Gabriela Goldschmidt**

Rachel Sebba**

---

* National Institute for Testing & Evaluation, Jerusalem, Israel
** Technion – Israel Institute of Technology, Haifa, Israel

Abstract

This study provides a panoramic evaluation of the selection system used by the Faculty of Architecture at the Technion, Israel. For several decades now, the Architecture Selection Test (AST) has been a central component of this system. This practical, open-ended test (in which planning, designing, presenting and analyzing abilities are tapped) is used in addition to high school matriculation (Bagrut) exams and to the all-university Psychometric Entrance Test (PET). The report starts with an introductory review of the selection procedures used in architectural departments around the world, followed by a description of the process of selection to the Tecnion's architecture school. Then the results of a brief job analysis are presented, or rather a personality analysis, portraying the qualities perceived as most desirable for the architecture student. Finally a predictive validity study is presented regarding the admission process employed by the Technion's architecture school, including a reliabilty assessment and a generalizability analysis of the AST. Our main conclusion is that even before implementing the recommended improvements in the processes of developing and scoring AST, its contribution to the prediction of the criteria it was meant to, namely, the performance in the studio (design projects) in architecture studies, can be clearly shown.

1. Introduction

Most departments in Israel's institutions for higher education use as their main selection tools high school matriculation exams and the national Psychometric Entrance Test (PET). The Faculty of Architecture at the Technion has employed, for many years, an additional test of special-abilities – the Architecture Selection Test (AST), intended to measure unique skills required for architecture studies, skills which are not measured by high school matriculation exams (Bagrut) and PET.

AST is a costly test to develop and administer, and especially to rate and score. Moreover, its contribution to the validity of the selection process has been frequently questioned. This evaluation study was aimed to examine the current selection procedure and to consider ways to increase its validity and decrease its costs.

The first step in this project was to conduct a review of the selection procedures in schools of architecture worldwide.

1.1 International Survey[1]

184 letters were sent in July 2000 to schools of architecture, requesting information about practices, policies and procedures pertaining to admission of students, encouraging also commentary on the subject. The mailing list was derived from the web site of the American-based Association of Collegiate Schools of Architecture (ACSA) and included full and affiliate member schools around the globe[2]. A reminder was mailed in December 2000, resulting in a total of 80 responses. Of those 70 are relevant to the survey; the rest include no relevant information or belong to categories of schools that are not compatible with the study's objectives (e.g., programs with graduate studies only). All 70 schools, in 22 countries, offer professional degrees in architecture; the programs are varied in their structure and duration of studies (between 4 and 6 years), as are the degrees awarded.

The survey revealed that there is a great diversity in admission philosophies, and widely held views regarding admission requirements are not necessarily anchored in reality. Lewis (1985/98), for example, claims that "most schools demand similar kinds of academic experience, knowledge, and abilities" (p. 123). This turns out to be imprecise. Moreover, there is little certainty within schools regarding the application of admission criteria. As a spokesman for a well known American school stated: "…Having taught entering students, along with graduate studies, for more than twenty years, it is always difficult to develop criteria as to who should study architecture."

We have extracted an exhaustive list of eight different criteria for admission into schools of architecture. All the schools in our survey use one or more of these criteria to screen candidates; no school admits applicants without any scrutiny. In some cases screening according to some of the criteria is not implemented to determine intake into first year studies; rather, it serves to control admission to a subsequent stage of studies. The criteria are general and each includes many variations that cannot be reported here in detail. We distinguish between two classes of admission criteria: general, i.e., criteria similar to those used in other academic disciplines, and special, i.e., criteria specifically relevant to design education (and related fields, like art).

[1] See detailed reports in Goldschmidt et. al. (2001, Forthcoming). Section 1.1, including Table 1, are reproduced from the forthcoming report.

Table 1 presents the frequency distribution of criteria.

Table 1. Percentage of schools using general and special admission criteria

General Criteria	% Use	Special Criteria	% Use
High school achievements	91%	Architecture aptitude test	27%
Scholastic aptitude test	56%	Portfolio	20/31%*
Interview	26%		
Personal statement			
Recommendations	17%		
Writing skills	14%		

* Obligatory/optional submission requirement

General Criteria

High school records: According to the norms in each country, high school records may consist of average grades such as the GPA (Grade Point Average) in the USA, a state-administered final exam such as the Matriculation, or a Baccalaureate (there even is an International Baccalaureate certificate).

Psychometric tests/General scholastic aptitude: In many countries university entry applications require that the candidate submit records of a general scholastic test and/or psychometric test. Such tests examine various cognitive and scholastic abilities to estimate future success in academic studies. These tests are normally administered by the state or by the universities; examples are the SAT (Scholastic Assessment Test) and ACT (American College Tests) systems in the USA.

Interview: Interviews of candidates by faculty members (and sometimes also by advanced students) are conducted in several schools. The weight of the interview results varies largely among schools. The interview can also include the presentation of a portfolio of creative/design work. In some cases interviews are not held at the point of entry into the school, but as part of the process of controlling admission into a higher phase of studies.

Essay: A few schools require essays – for the most part short ones (approximately 500 words); in other cases a longer writing assignment is given. The purpose is to test the candidate's ability to clearly communicate ideas and reason them through.

Written statement: Some schools require personal statements explaining why the candidate wishes to study architecture.

[2] Other sources of information were attempted, including the European Association for Architectural Education (EAAE), but no lists of schools were available on the web at that time.

Letters of recommendation: Letters of recommendation from former or present teachers or persons who are acquainted with the candidate's work and personality may also be required, often in conjunction with a personal statement.

Special Criteria

Special architecture aptitude test: Some schools use special tests that are believed to reveal candidates' aptitude for architectural studies. Tasks given in these tests are intended to examine, among others, spatial perception and organization, drawing skills, environmental sensibilities, and more. In most cases these tests are administered by the Architecture Departments themselves.

Portfolio: A portfolio of design work (where applicable) or other creative work is reviewed in quite a number of schools as part of the requirements for admission into first year studies or a subsequent year. In some instances a portfolio is required at more than one point along the way to graduation. A portfolio may also be required as part of an interview. In some schools the submission of a portfolio is voluntary and is encouraged in borderline cases.

As is evident from Table 1, some admission criteria are more prevalent than others. High school records figure in almost every case (91%). A good standing in psychometric/scholastic aptitude tests is required by approximately half the schools (56%) we have surveyed. Less widespread are interviews (26%), essays (14%), personal statements (10%) and letters of recommendation (13%) (when consolidated the latter two criteria amount to 17%). Special architectural aptitude exams are held by a quarter of the schools (27%) (with a single exception, none are held in North America). Portfolios are required in a smaller number of schools (20%), but are optional in yet another 11%. However, portfolios are often a requirement for subsequent advancement into a higher phase of studies (e.g., second or third year). The average number of criteria used for admission into first year studies, across schools, is 2.5.

Not much is known about the success of prevalent admission criteria to provide schools of architecture with the students they desire to enroll. Even less is known about the impact of selection methods on the built environment everywhere in the world. It is our hope that further studies will shed much needed light onto these important issues. The rest of this study is an attempt to do so.

1.2 The selection process at the School of Architecture at the Technion

Admission requirements[3]

For practically all applicants (excluding those with an academic background, or other special status), a special composite score is calculated. This score is supposed to reflect prospects of success, and by means of this score candidates are ranked and selected. This score is called the Architecture Composite Score (Arch. Comp.) and is based on:

Bagrut (Matriculation) average score – based on a weighted combination of high school grades and the higher of the following two: the average scores on national tests in various subjects, or, for applicants who undertake the full Technion's preparatory program, the program average scores. This score was allotted a weight of 37%.

PET (Psychometric Entrance Test) general score – allotted a weight of 30%. A total score in which PET's three sub-tests relative weights were: 40% Verbal Reasoning, 40% Quantitative Reasoning, and 20% English. These sub-tests are described below. PET is a multiple-choice battery, designed to measure various aspects of developed scholastic abilities and skills and aimed to predict future academic performance in higher education.The three sub-tests which comprise the PET are:

V (Verbal Reasoning) – This section includes 60 items that focus on verbal skills and abilities manifested in academic performance: analyzing and understanding complex written material, thinking systematically and logically, making fine distinctions among the meanings of words and concepts. The Verbal section includes items such as synonyms and antonyms, analogies, sentence completions, logic and reading comprehension.

Q (Quantitative Reasoning) – Includes 50 items that focus on the use of numbers and mathematical concepts (algebraic and geometrical) to solve quantitative problems and to analyze information presented in graphs, tables, and charts. In this sub-test, only a basic level of mathematics is involved – that which is acquired in the ninth or tenth grades in most high schools in Israel. Formulae and explanations of mathematical terms that may be needed in the course of the test are offered in the test booklet.

E (English as a foreign language) – Includes 54 items designed to test mastery of reading and understanding academic-level texts in English. This section includes three types of items: sentence completion, restatements, and reading comprehension. This sub-test serves a dual purpose: it is a component of the PET total score, and it is also used for placement of students in remedial English classes.

AST (Architecture Selection Test) – 33%.

In the studied period, 350 to 700 candidates have been administered the AST annually and up to 100 were admitted (see an example of the test in Appendix 1).

Purpose. According to the Technion, the purpose of this test is to ensure that the students admitted to the Faculty of Architecture have the skills needed for success in the projects and in the practical (technical/artistic) professions, that are manifested mainly in a figural way of expression and are acquired through experiencing creative processes, while implementing and integrating knowledge from diverse fields. The test was not intended to be a predictor of success in theoretical studies.

Structure. The time allotted for the test is five hours (including one half-hour break). It includes three sections. In each section the examinees are required to choose one of the two tasks presented, so they have to perform three tasks in total[4]. Two sections are introduced at the beginning of the test and the third is introduced after the intermission. In their answers the examinees are expected to produce graphic descriptions and verbal explanations for the objects they plan, design or draw from memory.

Criteria for evaluation. In April 2000 a detailed list of the criteria used in evaluating the answers to the test's tasks was compiled. These criteria were based on guidelines which appeared in a Technion report (Sebba, 1989). In the past, however, the evaluation of the test was not based on a defined list of criteria.

In Appendix 2 these criteria are presented with an accompanying explanation for each criterion. Reading this list of criteria, one cannot easily distinguish between **traits** sought for, and **operational manifestations** of traits.

---

[3] It should be noted that these requirements hold for the populations under study, namely, applicants from the years 1991-1999, and may have since slightly been modified, especially since 2002.

[4] In March 2000 the test was shortened so that it would include two sections only, in order to reduce the burden placed on the applicants. The time allotted for the test was not changed. In March 2001 the format of three sections was reinstated, but only one, mandatory task is presented in each section (see section 3.2). In this study we will refer to the three section format.

It is highly recommended that a well-defined list of specifications and a scoring key be developed to be used by the test evaluators. This should be done by first, finding out the seven or so traits that are most accepted as important for the architecture student. These traits should be defined as clearly as possible and should be distinguishable from one another as possible. Each task in the test should tap some of these traits explicitly, so that the test as a whole will examine them all. Each trait will then be translated into one or more operational manifestations, which are less abstract, and more direct pieces of evidence to be observed in the test. The evaluator will be instructed to assign any such evidence a numeric value, which is proportional to the degree to which this trait is manifested in the solution to the task. The average of these scores will be transformed into a numerical score of the trait to which they attest. A profile of the scores of the primary seven or so traits will then be available, and the sum of these sub-scores will conclude the examinee's final score. The first step toward the development of this procedure and an example are presented in the following personality-analysis chapter.

Evaluation[5]. The test evaluation was performed in a three-step, rather hirearchical process, by a few faculty members. First, the works (bearing codes only) were assigned by teaching assistants into one of ten levels of quality. Then, two senior faculty would approve the primary quality allocation, making occasional relocations if necessary, and consulting one another in doing so. Finally, an expert teacher would grade each work numerically, according to the pile in which it was found.

We recommend that each section of the test be evaluated independently of the others, by two or more raters, who will assign it a numerical score. If the difference between two scores is large (the acceptable difference between scores will be predetermined), this section is to be rated by yet another judge. The final score would be the average of the scores assigned by the raters[6]*. A well-constructed rating workshop administered to the raters, along with a*

[5] The investigation undertaken in this study regarding the predictive validity of the Architecture Selection Test made use of data collected before 2001. For this reason we describe the evaluation procedure which was used up to this year. In 2001 a new procedure was administered, which will be described in the section regarding the Generalizability analysis.

[6] A study recently conducted at NITE (in preparation) suggests that in a case of a large gap between two evaluators, adding a third is of little or no use. If one does insist to add a third evaluator, a better estimate of the true score is achieved by averaging the three evaluations than by replacing the extreme evaluation with the third one.

detailed, precise scoring guide, seem to be two necessary measures to be taken in order to ensure reasonable inter-rater reliabilty.

2. Who should study architecture: A personality profile

We tried to determine which traits faculty members and outside architects regard as important for an architectural student. The data were collected both directly – by asking teachers to fill out a questionnaire[7], and indirectly – performing a post-hoc analysis of the wording and the argumentation accompanying the nomination of prize-winning architectural projects. A third source for this inquiry was the AST rating specifications (Appendix 2), which has been in use for the last three years or so, and reflects predominant ideas in the Technion Faculty of Architecture.

2.1 The Teachers Questionnaire

Thirteen senior faculty members responded to a questionnaire (constructed by a senior professor in the Faculty in consultation with peers) – in which they were requested to rank each trait (out of a list of fifteen) on a scale from 1 to 5 (on which 1 designates "extremely important", 2 - "very desirable", 3 - "desirable", 4 - "unimportant" and 5 "interfering with") – according to their degree of importance in characterizing students in (a) early stages and (b) advanced stages of study. In an open-ended question, the teachers were also encouraged to add important traits that were not included in the list of fifteen.

Table 2 lists the traits offered in the questionnaire, ordered according to their perceived degree of importance. The degree of conformity among teachers can be inferred from the s.d. values.

[7] A similar test was conducted in the past at the Technion (see Sebba, 1989).

Table 2: Ranking of traits from the teachers' questionnaire

(low average reflects high rating of importance)

	Early stages of study			**Advanced stages of study**		
	Trait	**Ave.**	**s.d.**	**Trait**	**Ave.**	**s.d.**
1	Imagination	1.46	0.66	Abstract thinking	1.38	0.51
2	Creativity	1.67	0.71	Creativity	1.50	0.71
3	Originality	1.67	0.78	Originality	1.58	0.79
4	Abstract thinking	1.69	0.75	Imagination	1.62	0.51
5	Conceptual flexibility	1.75	0.62	Environmental sensitivity	1.67	0.65
6	Environmental sensitivity	2.08	0.90	Conceptual flexibility	1.75	0.45
7	Figural literacy	2.09	1.14	Delay of judgment	1.75	0.62
8	Delay of judgment	2.25	0.75	Technical reasoning	1.75	0.62
9	Social sensitivity	2.50	0.90	Figural literacy	1.91	0.83
10	Organizational skills	2.62	0.65	Perception of details	2.00	0.60
11	Technical reasoning	2.67	0.49	Organizational skills	2.08	0.51
12	Visual memory	2.67	0.65	Social sensitivity	2.08	0.79
13	Associative thinking	2.67	1.07	Associative thinking	2.42	1.00
14	Perception of details	2.85	0.80	Visual memory	2.55	0.69
15	Preference for complexity	3.11	0.78	Preference for complexity	3.25	1.04

A comparison between the two rankings – for the early and for the advanced stages of study - shows a remarkable concordance between them. In general, the traits receiving the highest ratings are: imagination, creativity, originality, abstract thinking, conceptual flexibility and environmental sensitivity. The lowest average rating of a trait was 3.25; that is, all traits were rated, on average, from desirable to extremely important. It should be noted that on average, most traits (except preference for complexity) are perceived as more important in advanced phases of study than in early ones.

The following is a list of the qualities suggested by the teachers as highly important, in the open-ended section of the questionnaire (we are aware that due to the small size of the sample, their value is mainly impressionistic, though rather enlightening): motivation to excel, reexamination of conventions, involvement in contemporary culture, conceptual openness, independence, general knowledge, diligence and perseverance, emotional maturity, ability to make decisions, written and oral expression, graphic expression, emotional intelligence, spatial perception.

2.2 Evaluation of prize-winning projects

Fifty-one evaluations of twenty-seven design projects performed by students at the School of Architecture and chosen as outstanding in contests held in the Faculty were collected. The evaluations, submitted by Faculty members from the faculty and by outside architects, included the argumentation and rationale for choosing the project as outstanding. The wording of these evaluations were analyzed in order to extract the traits reflected in them, which led them to win the prize. As far as possible this verbal analysis was made using the traits listed in the teachers' questionnaire. This was done in order to facilitate a comparison between these two sources of information. Table 3 lists the traits, ordered according to their frequency of use:

Table 3: Traits extracted from 51 evaluations of prize-winning works

Trait	Frequency of use
Environmental sensitivity	28
Originality	26
Figural literacy	13
Profundity	11
Technical reasoning	10
Problem solving ability	10
Clarity	9
Functionality	7
Creativity	7
Analytical skills	4
Cultural sensitivity	3
Graphic expression	3
Complexity	3

The two most frequently mentioned traits in the evaluations of the outstanding projects are: environmental sensitivity and originality. In the list of the highest seven traits, we also find: figural literacy, profoundity, technical reasoning, problem solving ability, and clarity.

A brief synthesis of all three sources (the teachers questionnaire, the prize-winning evaluations and the AST Rating Specifications - Appendix 2) shows that among the top ten or so most desirable traits, all three sources share agreement about the following common qualities: creativity, originality and imaginativeness (as closely related), figural/spatial literacy/reasoning, technical reasoning, problem solving ability, environmental sensitivity. In addition, each source would add some unique traits, such as verbal ability, visual memory, preference for complexity and speed of work – by the AST Rating Specifications form; abstract thinking, organization ability, delay of judgement and conceptual flexibility – by the teachers questionnaire; and depth, clarity and functionality – by the prize-winning judgements.

We think that the aforementioned list of traits, especially with respect to those agreed upon by all sources, is a good start in determining what an architecture special ability test should measure. The next challenge for test assemblers, and not a simple one, would be the operationalization of these qualities into measurable assignments.

3. The predictive validity of the selection procedure

Preface

As described above, the selection system under study comprises three ingredients: Two scholastic, general ability measures – one which is more curriculum-oriented, the *Bagrut*, and one which is less so – PET. In addition, it includes a special ability measure for architecture – the AST. The main purpose of this part of the study is to assess whether the AST is a valid tool in predicting external criteria of success in practical projects (in design- and planning-oriented courses), and if so, does it contribute to the validity of the selection procedure beyond that of the scholastic aptitude components: PET and *Bagrut*. An interesting by-product of the study would be, of course - in addition to the estimate of the validity of the whole process – a measure of the relative importance of all the components of the selection procedure: AST, *Bagrut*, PET and its three components.

Some School of Architecture faculty continue to express dissatisfaction with AST's costliness, while questioning its predictive benefit. Previous validity studies concerning this selection system (Sebba ,1989, and Technion Working Paper (1997)) report "medium to low validity" of AST (raw correlations of around 0.30 or less). A more appropriate body of predictive-validity data for reference, in terms of breadth and methodology (although not concerning AST), is the meta-analytic study by Kennet-Cohen et al. (1999a). The validities reported (averaged across 161,000 students, from the classes of 1984 to 1995 of all Israeli universities, altogether 1,861 units of analysis) are corrected both for restriction of range and criterion (FY scores) reliability. The estimated true validities are: PET– 0.45, Bagrut – 0.41, Composite score – 0.55.

Method

Population: 547 Technion students from nine architecture classes of the years 1991 through 1999, for whom full records were available (of the Bagrut, PET, AST and first year grades, in both theoretical courses and design projects[8]).

Criteria. The criteria for success in school were first-year average[9] scores. Three types of criteria were defined:

1) FY average score in practical design projects (Project score).
2) FY average score in theoretical courses (Course score).
3) FY average of scores in both courses and projects (Combined score).

Reliability of the criteria. Since the inner consistency of a criterion creates an upper limit to its expected predictability, we questioned the reliability of the criteria used in this study. Of major interest was the reliability of the design-project evaluations, used for the validation of the AST, as these evaluations are evidently and admittedly "soft" and not very well-structured by nature.

The estimation was conducted as follows: All semesterial scores of each student were randomly divided into two groups (split-halves). Each subject contributed to each half an average semesterial score. A Pearson correlation coefficient between the the two averages was calculated. Then, a correction for split-half reliability was implementedaccording to the Spearman-Brown formula[10].

This calculation shows, not surprisingly, that the estimared reliability of the FY Project score - 0.56, is lower than that of the FY Course score - 0.59, and the estimated reliability of the FY Combined score - 0.66, is higher than that of each of its two components.

Predictors. The following variables served as predictors (fully described above in section 1.2):

1) **AST** (Architecture Selection Test) score.
2) **Bagrut** (Matriculation) average.
3) **PET** (Psychometric Entrance Test) general score.

[8] Grades of courses from two semesters of study, the credits for which totalled 12 or more.
[9] All avereges were weighted by number of credit points of the projects or the theoretical courses.
[10] For detailed description, see Appendix 3.

4) **V** (Verbal Reasoning).
5) **Q** (Quantitative Reasoning).
6) **E** (English as a Foreign Language).
7) **General Composite** – Total score composed of PET and Bagrut, equally weighted[11].
8) **Architecture Composite** – Total score composed of PET, Bagrut and AST, equally weighted[12].

Unit of Analysis. The basic unit of analysis in this study is a single architecture class, of about 60 students on average. All statistics presented are weighted averages (by class size) across all classes.

Results

Descriptive statistics. Table 4 presents the averages and standard deviations of the predictors and the criteria.

Table 4: Averages (and standard deviations) of FY scores and predictors

	Predictors								*Criteria*		
	AST	Bagrut	PET	E	V	Q	General Comp.	Arch. Comp.	Course	Project	Comb.
Ave.	83.87	100.76	681.58	132.81	130.16	132.76	82.55	83.10	84.70	82.75	84.09
S. D.	(9.28)	(6.36)	(39.56)	(12.16)	(10.20)	(8.80)	(4.81)	(3.10)	(4.38)	(6.02)	(3.87)

Validities. Pearson correlation coefficients between the predictors and the criteria are presented in table 5. The values are corrected for restriction of range[13]. The correction formula (Gulliksen, 1950, see Appendix 5) assumes that the restriction of variance in each predictor is, in fact, indirect, and is an indirect result of the restriction by the (direct) selecting tool - Arch. Comp. The unrestricted standard deviations were estimated from the data of applicants to each class.

[11] The formula for calculating this score was not identical for all classes. Therefore, we employed uniformly the formula (see Appendix 4) used in the latet two classes. This formula produced a composite score with an average of 82.55 and a standard deviation of 4.81.
[12] The formula for calculating this score was not identical for all classes Therefore, we used the Arch. Comp. calculated by the Technion (average 83.10, standard deviation 3.10). The General Comp. score was not included in the data provided by the Technion. Therefore, one should be careful when comparing General Comp. with Arch.Comp.
[13] Raw (observed) correlations are presented in Appendix 6.

Table 5: Validities corrected for restriction of range

	Predictors							
Criteria	AST	Bagrut	PET	E	V	Q	General Comp.	Arch. Comp.
Course	0.31	0.48	0.45	0.35	0.25	0.44	0.54	0.54
Project	0.29	0.18	0.19	0.15	0.10	0.20	0.22	0.30
Combined	0.36	0.44	0.42	0.33	0.22	0.42	0.50	0.54

The findings show that:

- The best single predictor of the Project scores is the AST (0.29). Adding the Bagrut and PET to AST (resulting in Arch. Comp.) hardly improves the Project's prediction (0.30).
- In predicting the Project score, Arch. Comp. shows a considerable advantage over Gen. Comp. (0.30 vs. 0.22, respectively).
- Arch. Comp. and Gen. Comp. show similar ability in predicting the Course score (0.54).
- In predicting the Combined score, Arch. Comp. shows a noticeable advantage over Gen. Comp. (0.54 vs. 0.50, respectively).
- As a whole, the Course scores prove more predictable than the Project scores. This pattern is still evident, if relatively reduced, after correcting the validities for differences between the reliabilities of these two criteria. If we apply the aforementioned findings, and correct our attenuated validities stemming from criterion reliability (divide by the square root of the estimated reliability, respectively), we obtain the following values:

Table 6: Validities corrected for both restriction of range and criterion reliability

	Predictors							
Criteria	AST	Bagrut	PET	E	V	Q	General Comp.	Arch. Comp.
Course	0.40	0.62	0.58	0.45	0.32	0.57	0.70	0.70
Project	0.39	0.24	0.25	0.20	0.13	0.27	0.29	0.40
Combined	0.44	0.54	0.52	0.41	0.27	0.52	0.62	0.67

It seems that this additional correction preserves the whole pattern of correlations, resulting in the same set of conclusions. From a meta-analytic point of view, these corrected values are even considered to be closer to the "true" values of validity.

***Estimating predictors' relative importance to prediction: Multiple regression*.**

The purpose of the following analyses was to estimate the relative weights of the predictors in predicting FY scores. These weights can be interpreted in terms of relative importance or additive contribution to the prediction. Two models were tested: a three-variable model: AST, *Bagrut*, and PET; and a five-variable model: AST, *Bagrut*, V, Q, and E. All three FY criteria served as independent variables.

In table 7, multiple correlations (R) and standardized regression weights (β's) are reported for the two models. These regressions are based on the correlation coefficients[14] corrected for restriction of range.

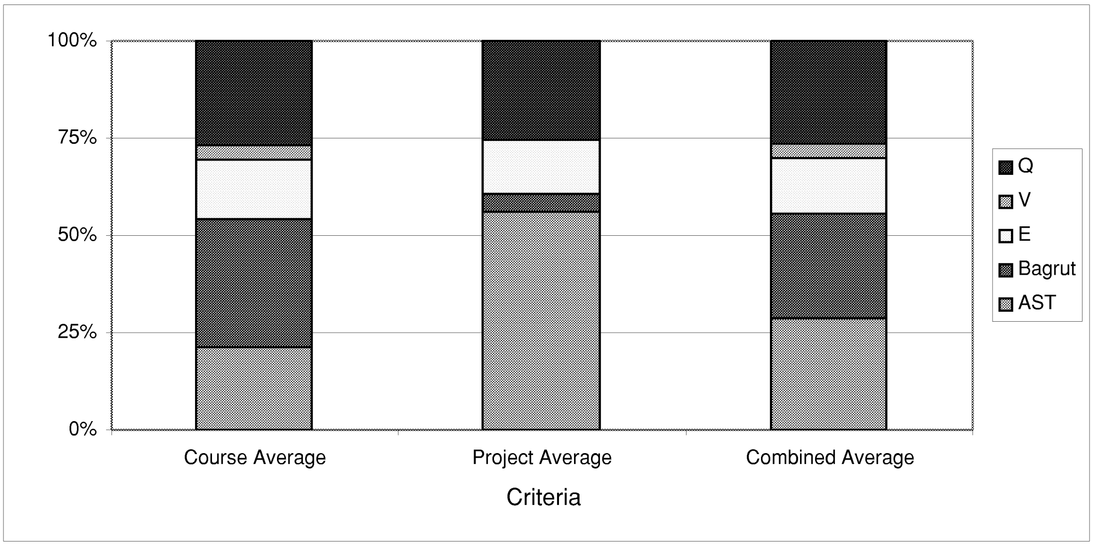
Table 7: Partial coefficients (β) and multiple correlations (R) in multiple regressions predicting FY scores

	Three-variable model				Five-variable model					
Independents:	AST	Bagrut	PET		AST	Bagrut	E	V	Q	
Dependents:	β			R	β					R
Course score	0.15	0.30	0.26	0.58	0.18	0.28	0.13	0.03	0.23	0.61
Project score	0.23	0.04	0.12	0.41	0.24	0.02	0.06	-0.00	0.11	0.44
Combined score	0.22	0.24	0.25	0.58	0.24	0.22	0.12	0.03	0.22	0.60

Chart 1 illustrates the marginal contribution (proportinal to their β's) of the five variables - AST, *Bagrut* and PET's components (V, Q, E) - predicting FY scores.

[14] The correction was done by performing the multiple regression analyses using the corrected correlation matrix of the variables included. The regression coefficients based on the raw correlation coefficients are presented in Appendix 7.

Chart 1: Relative weights[15] in predicting FY criteria



- When predicting the Combined score, the relative weights of the predictors are approximately as follows: AST – 30%, *Bagrut* – 25%, Q – 25%, E – 15%, and V – 5%.
- The relative weight of AST is about 55% in predicting Project scores, and about 20% in predicting Course scores.
- For each of the three criteria, the weight of PET components altogether is about 50%. Among the three components, Q seems to be the most important contributor (25%), while V hardly shows any meaningful additive contribution to the selection process.
- *Bagrut's* relative weight is about 35% in predicting Course scores, and only about 5% in predicting Project scores.

[15] The relative weights are the proportions of the sum of the β's.

AST is the best single predictor of the Project scores (in comparison to the *Bagrut* and PET). In predicting the Combined scores, the marginal validity of AST is substantive and in predicting the Project scores AST's marginal validity is very high. The results indicate the best weights to be assigned to each component of the selection system. As expected, the scholastic aptitude predictors (PET and Bagrut) are the best predictors of the theoretical Course scores (especially when the two are combined). AST also contributes to the prediction of the Combined scores beyond the contribution of PET and Bagrut. Therefore, it seems that even in its previous, as yet unimproved mode – of construction, of administration and of rating – AST provided a unique contribution to the selection procedure, in the domains for which it was meant to predict.

3.2 A Generalizability analysis

In the March 2001 administration of AST, several modifications were introduced to the procedure of scoring. Three independent raters evaluated the test. Each rater assigned a separate score to each of the three tasks chosen by the examinee, and then calculated an average score for the test. The final test score was the mean of the three average test scores produced by the three raters.

334 applicants were then administered the test.

Table 8 displays the average and table 9 – the standard deviations of scores by sections, tasks and raters[16].

Table 8: Average scores by sections, tasks and raters

Section	Task (Rater →)		A		B		C		All	
A	1	n=110	55.8	59.7	61.1	55.4	70.0	66.8	62.0	60.6
	2	n=224	61.6		52.6		65.3		59.7	
B	1	n=193	61.6	60.4	59.7	56.6	67.8	66.6	62.9	61.2
	2	n=141	58.7		52.3		64.9		58.4	
C	1	n=137	61.0	60.0	59.9	57.7	71.4	65.7	64.1	61.1
	2	n=197	59.3		56.2		61.8		59.0	
All		n=334		60.0		56.6		66.4		61.0

Table 9: Standard deviations of scores by sections, tasks and raters

[16] The tables include data supplied by the original three raters only.

Section	Rater / Task		A		B		C		All	
A	**1**	n=110	16.6	17.0	15.2	19.9	13.9	14.4	16.1	17.9
	2	n=224	17.3		21.4		14.4		18.6	
B	**1**	n=193	17.4	18.1	19.0	20.2	13.0	13.1	17.0	17.8
	2	n=141	18.9		21.1		13.7		18.7	
C	**1**	n=137	19.3	16.4	20.4	18.0	12.9	15.2	18.5	16.9
	2	n=197	14.2		16.0		15.4		15.4	
All		n=334		60.0		17.2		19.4		14.3

One can observe large differences in leniency between raters (a ten-point gap between raters C and B, rater A being in the middle). These differences show that either the raters were not trained and guided properly, and/or the evaluation guidelines were not precise enough. There are no large discrepencies between the sections' averages, but, over raters, tasks 2 "scored" lower than tasks 1 (differences ranging from 5.1 and 4.5 points in sections C and B, and 2.3 points in section A). With no further inquiry – in which, for instance, examinees would be randomly assigned to tasks, or all tasks would be administered to a group of examinees – it is difficult to explain these differences, which might derive from one or more of the following reasons: tasks 2 were more difficult than tasks 1; the examinees who chose tasks 2 were of lower ability; there were technical flaws or errors in the process of administration. In any case, these findings do not support the assumption of equivalence between the two tasks in each section (see footnote 4).

Tables 10 presents, for each rater, the Pearson correlations between the scores of the sections. Table 11 shows the correlations between the raters' average test scores.

Table 10: – correlations between sections within raters

Rater	Section	A	B
A	**B**	0.83	--
	C	0.73	0.79
B	**B**	0.87	--
	C	0.78	0.81
C	**B**	0.74	--
	C	0.61	0.65

Table 11: correlations between raters

Rater	A	B
B	0.73	--
C	0.68	0.67

The average correlations between the sections range from 0.82 for rater B, to 0.67 for rater C. These correlations might be somewhat spurious, as a result of dependency: each rater obtained all three tasks of an examinee together. In order to lower this artifactual dependency between tasks, it is recommended that every task of each examinee be distributed to a different rater.

The average correlation between raters is 0.69 (ranges from 0.67 to 0.73). This correlation is an indication of the between-rater reliability of the test. It is not particularly high, but it is probably an underestimation of the true reliability, because, as aforesaid, the examinees did not perform identical tasks.

Following is a more integrative estimation of the different aspects of reliability of the test, and an indication of the relative "share" each of the different participants (raters, tasks, subjects) have in the unexplained variance of test scores.

A generalization analysis was carried out on the data described above, based on Generalizability Theory (Brennan, 1983), and using GENOVA software v2.2 (Crick & Brennan, 1984). The purpose of this analysis was to estimate the contribution of the different variables to the error of measurement. The results of this analysis can also teach us about alternative test models, which include different combinations of number of tasks and number of raters. The generalizability analysis also yields an index of the reliability of the test in the way it was evaluated, where the higher the reliability, the higher the generalizability of the test. In the case of AST we are dealing with a two-factor design, where the possible sources of the score variance, beyond inter-personal ability differences, are: differences between raters, differences between tasks (the two principal factors), and interactions – examinee-rater, examinee-task, rater-task, and examinee-rater-task. However, the actual test administration design, in which each examinee was required to choose one of the two tasks in each of the three sections of the test, limits the extent of generalizability of the findings. Methodologically speaking, we would have been better off had all examinees responded to all tasks. Specifically, it is not evident that the two tasks in each pair are equivalent, both regarding their difficulty and the constructs they measure. For practical purposes – with three tasks for each examinee, where each task was evaluated by all three raters, and forcing the assumption of equivalence of these pairs of tasks within sections, and assuming randomness of choosing them – a generalizabity analysis was performed. The reported design is crossed, namely, it assumes that all raters evaluated all tasks performed by all examinees (this is only an approximation because, as aforementioned, in each section of the test the examinee could choose one of two tasks; see footnote 4). The results of the generalizability analysis for the 334 observations are presented in table 12.

Table 12: Generalizability analysis*.

Effect (source of "noise")	(1) Variance components for single observations	(2) Estimates of variance components for average scores	(3) Standard errors
Examinees P	158.92	158.92	14.93
Raters R	24.33	8.11	5.84
Tasks T	0.00	0.00	0.05
Interaction PR	64.91	21.64	1.50
Interaction PT	17.78	5.93	0.66
Interaction RT	0.64	0.07	0.05
Interaction PRT	49.84	5.54	0.21
Generalizability coefficient: 0.83			

* The values in column (1) are the values of the variance components for single observations, with a generalizability ability for an infinite population (identical in a crossed design for D and G studies). In column (2) the generalizability power of the estimates for the variance components (in a D study) is for our specific case of the studied design, with the three specific raters employed and the "three" tasks included in the test.

For the specific model studied, the generalizability coefficient, which can also be interpreted as the reliability coefficient, is 0.83. This is a relatively high value for a test of this sort – a performance assessment test. The variance components that do not originate from the examinees (P), can be interpreted as "noise", or the unaccounted-for variance.

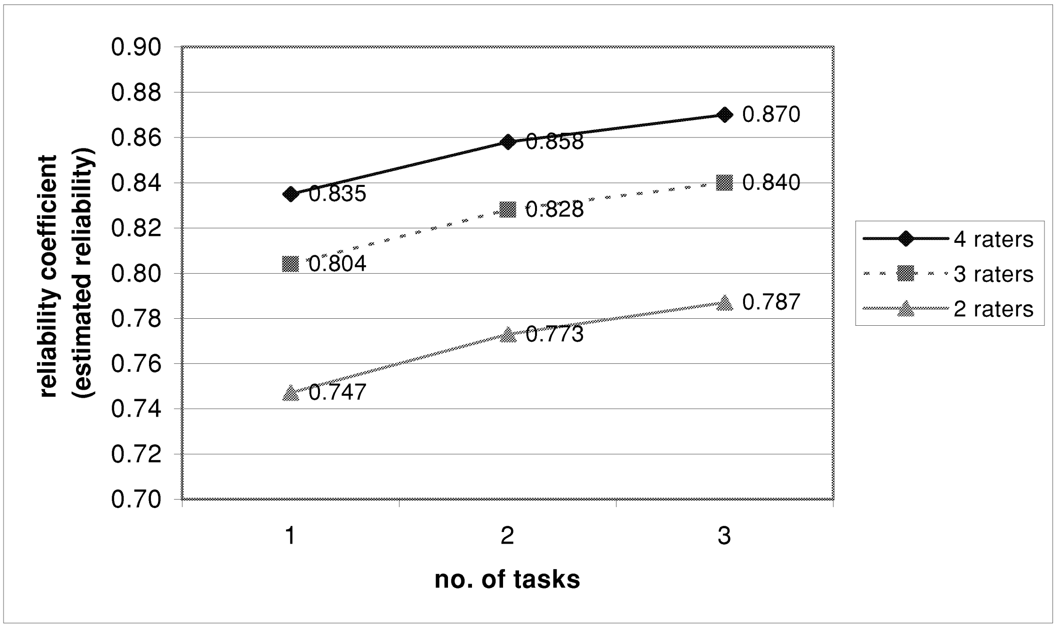
The raters stand out as a major main-effect source of "noise", and even more so – the interaction between raters and examinees. The meaning of this interaction is that the examinees are rated differently by different raters. This takes us back to the notion that the raters were not trained in a satisfactory way, or that the criteria for evaluation were not precise and well-constructed enough, or both. Proper training is supposed to reduce the idiosyncratic nature of this evaluation.

On the other hand, the variance of the tasks do not prove to be a source of the unaccounted-for variability.

In any case, even with other things kept equal, reducing the noise (or increasing the reliability) that stems from these effects is possible both by increasing the number of tasks, and by increasing the number of raters.

The following analysis shows the amount of gain or loss, in terms of change in the reliability of the test, by modifying the number of tasks and raters. The analysis is in a crossed design in a D study which assumes raters and tasks of the sort that participated in the actual AST studied.

Chart 2: Estimated reliability of AST as a function of number of tasks and raters



The estimate of the reliability of AST (the way it was administered in 2001) is 0.83. If the test's cost is too high, the number of tasks or the number of raters should be reduced. Reducing the number of tasks to two will yield a test with a reliability of 0.80. Reducing the number of raters to two will reduce the reliability to 0.77. It seems that in the specific case studied, raters account more for AST reliability than tasks do.

4. Discussion

Compared to other selection systems in the world, it appears that the Technion School of Architecture operates a relatively rich selection system. Using scholastic aptitude tests, which include both school achievement (*Bagrut*) and academic potential tests (PET), together with a specific selection test (AST) comprising tasks which are believed authentic and relevant to the ones the student will be faced with during his/her studies and even during his/her professional work, facilitates a solution to the selection problem, or at least permits an appropriate choice among the selection system's components. It may still be possible to improve the specific architecture selection test, as was mentioned earlier.

As agreed upon by three sources of data, the desired architectural student is expected to possess high values of the following most important traits: creativity, originality and imaginativeness (as closely related), figural/spatial literacy/reasoning, technical reasoning, problem solving ability, environmental sensitivity. In addition, each source added some unique traits, such as verbal ability, visual memory, preference for complexity and speed of work, abstract thinking, organization ability, conceptual flexibility, depth, clarity and fuctionality. It seems that at least the first list (of consensual traits) provides a good start to test designers and assemblers.

Testing the predictive validity of the selection system components reveals, in terms of "true" validities (corrected for both restriction of range and criteria reliability), the following picture: AST stands out as the best single predictor of FY Project scores (0.39), and its marginal validity, beyond that of the other two scholastic components, is high (0.40 vs. 0.29). In predicting the Combined FY score, the marginal validity of AST is substantive (0.67 vs. 0.62). Relative weights of the components are suggested for optimizing the selection process. When the criterion of success chosen is FY Combined score, the equal-weights scheme for AST, Bagrut and PET seem to work well. In studying the improved AST version of 2001, both differences between difficulty level of tasks within pairs, and between the degree of leniency among the three raters, were found to be considerable. The Generalizability Analysis shows that in general the reliability of AST is relatively high (0.83). It was recommended not to allow a choice between tasks and this recommendation has indeed been implemented as of 2002 (see footnote 4). Also, it was recommended to provide thorough training to the raters along with clear evaluation instructions, in order to lower the variance among them, and to reduce the strong effect of rater/examinee interaction (dependency) detected. The difference between the raters proved to be larger than the difference between the tasks; therefore, it seems that in order to augment the reliability of the test, adding a judge would seem preferable to adding a task.

Comparing our results with other validity studies, we conclude that AST, from a validity point of view (both criterion-based and face validity) is a justifiable screening tool for schools of architecture which are interested in the design-oriented skills of its students, and much less so – when their focus lies mainly in academic skills.

Natural suggestions for future research would be to implement our recommendations for improving both the test and its scoring procedure, conduct a follow-up study to see if and how much the validity changed, and if possible add advanced-years performance indices as validating criteria.

References

Brennan, R. L. (1983). *Elements of Generalizability Theory*. The American College Testing Program. Iowa City.

Crick, J. E. & Brennan, R. L. (1984). GENOVA version 2.2.

Goldschmidt, G., Sebba, R., Oren, C. & Cohen, A. (2001). Who Should Be a Designer? Controlling Admission into Schools of Architecture. *Proceedings, 5th International Design Thinking Research Symposium: Designing in Context*. Delft University of Technology, December 18-20, 2001, pp. 277-295.

Goldschmidt, G., Sebba, R., Oren, C. & Cohen, A. (forthcoming). Crossing the Threshold: Criteria of Admission into Schools of Architecture.

Gulliksen, H. (1950). Theory of mental tests. New York: John Wiley & Sons. {Reprinted in 1987. Hillsdale, N.J.: Lawrence Erlbaum}.

Kennet-Cohen T., Bronner S., Oren C. (1999a). A meta-analysis of the predictive validity of the selection process used by universities in Israel (in Hebrew). *Megamot, Behavioral Sciences Journal* 40 (1) 54-71.

Kennet-Cohen T., Bronner S., Oren C. (1999b). The Predictive Validity of the Components of the Process of Selection of Candidates for Higher Education in Israel. *NITE Research Report No. 264*, p. 20. Jerusalem.

Lewis, R. K. (1985/1998). Architect? A candid guide to the profession. Cambridge, MA: MIT Press.

Linn, R. L. (1983). Peaarson selection formulas: Implications for studies of predictive bias and estimates of educational effects in selected samples. *Journal of Educational Measurement, 20*, 1-15.

Feldt, L. S., & Brennan, R. L. (1989). Reliability. In Linn, R. L. (Ed.), *Educational Measurement* (p. 117). Macmillan Publishing Company, NY.

Sebba, R. (1989). [*The Process of Selection of Applicants to the Faculty of Architecture and Town Planning at the Technion – Background, Criteria, Validity, Implications*. (in Hebrew)]. The Faculty of Architecture and Town Planning, The Technion – Israel Institute of Technology, Haifa.

The Technion – Israel Institute of Technology (1997). The Relation Between Admission Scores and Academic Success in The Faculty of Architecture and Town Planning. *Working Paper*. Haifa.

Appendix 1

Architecture Selection Test

26.3.1999

Instructions

The test is composed of three parts. One question in each part must be answered. Altogether three questions must be answered. All questions are granted an equal number of points. The total time allotted is four and a half hours. You should allot an hour and a half for each question.

Part 1	Part 2	Intermission	Part 3
One and a half hours	One and a half hours	Half an hour	One and a half hours

Read all the questions and instructions carefully before you begin.

* The questions should be answered using drawings, illustrations and sketches and should be accompanied by short and clear written explanations, on the supplied sheets of paper. Each answer should be complete and contain a short graphic description while the written explanations should be as short as possible.
* It is allowed and even recommended that the illustrations and sketches be done by hand only.

  You are permitted to use a pencil and the drafting tools you were requested to bring.
* Try to present your answers in a clear and methodical manner.
* The thick sheets of paper are designated for the test. The thin sheets of paper are designated for the drafts.
* At the end of the test return all the sheets of paper you were given (including the drafts and the test form).
* Ensure that your examinee number appears on every page of the test and that all parts of the test are stapled together when the test is over.

Good Luck!

Section 1

In this section there are two questions of equal value. Answer only one of them. You must answer each part of the question you have chosen. Please read the whole question carefully before you begin. It is recommended that you organize the answer in a methodical, efficient and concise manner.
The allotted time for the answer: one and a half hours.

1-A. An entrance

When we arrive at a certain place we reach it through an entrance. Entrances to a house, an apartment building, a theatre, a park and a street, are examples derived from private or public domains and can be constructed or open. Entrances are different from one another in their function, components, shape and size.

1.1 Using drawings, give five examples of different entrances.

1.2 Explain the function of each of the entrances you have chosen. Describe its components and the connection between its function, structure and form. It is permitted to add an explanation of other characteristics that you consider important.

1.3 As a summary, explain the important properties that one should consider when designing an entrance.

1-B. Chairs

Chairs are used for sitting in open or closed places. In different ages and cultures chairs and benches were suited to special conditions and local needs.

2.1 Draw five examples of different chairs (or seats).

2.2 Offer a <u>brief</u> explanation of the function of each of the chairs (or seats) you have chosen. Describe the materials each is made of and the connection between its function, material and shape. It is permitted to add an explanation of other characteristics that you consider important.

2.3 Make one of the chairs you have chosen suitable for [the use of] a one year old child and describe the new chair using a drawing or sketch and include a short written explanation.

2.4 As a summary, explain the common components of all the chairs and what can be learned about the society using them.

Section 2

In this section there are two questions of equal value. Answer only one of them. You must answer each part of the question you have chosen. Please read the whole question carefully before you begin. It is recommended that you organize the answer in a methodical, efficient and concise manner.
The allotted time for the answer: one and a half hours.

2-A. A childhood place

As the years go by, a special place where we grew up becomes a part of us. Demonstrate a place (inside or outside) that you remember as meaningful in shaping your personality and identity.

1.1 Describe the place using illustrations and/or a perspective drawing.

1.2 Give a written explanation of the connection between the properties of the place and its influence on you.

2-B. A resort

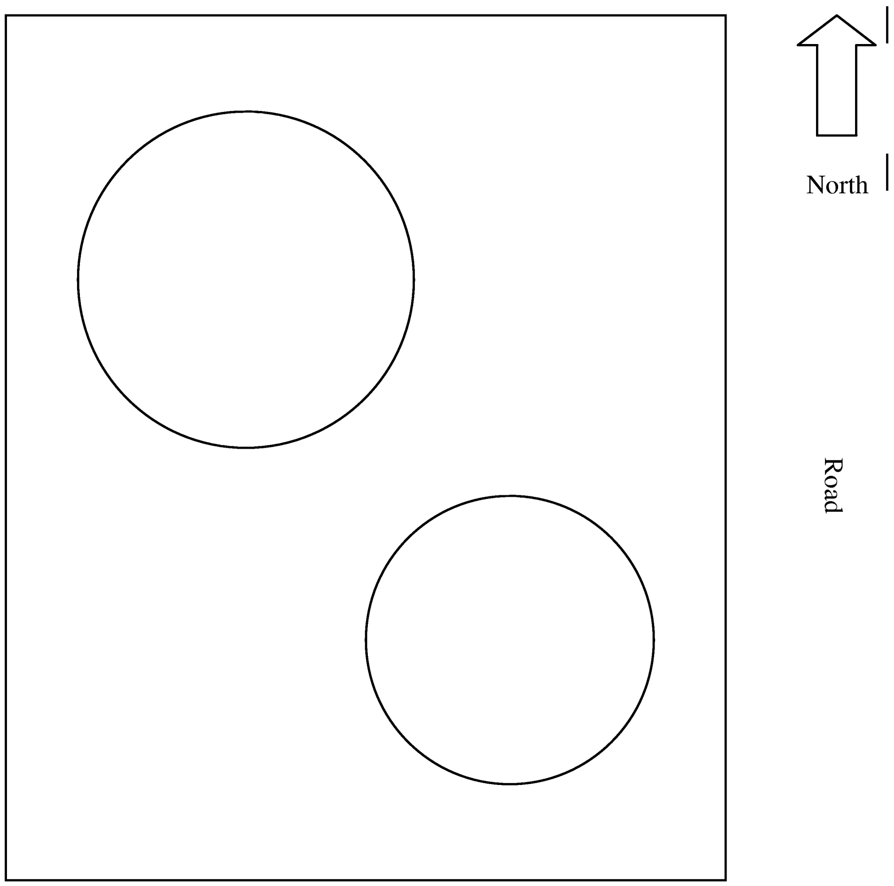
A tourism firm bought a lot with industrial buildings on it by the sea and wishes to develop it into a resort. On this land there are two round buildings (cylinders) which the firm wishes to transform into resort homes.
The abode is intended for the use of one or two families, according to your choice.
Design the Resort:

1.1 Design the entrance to the resort from the road area.
2.1 Show how the yard surrounding the houses will look like.
2.2 Plan the internal partition of the buildings and the location of the entrances.

Describe the spaces you have created (inside and outside) using drawings and/or illustrations and give a brief description of your design considerations.

Comment: Pay attention to the following data:
The diameter of the large building is about 8.00 meters. The diameter of the smaller building is about 7.00 meters. The buildings' height is 3.5 meters. The design instructions do not allow for the enlargement of the buildings or changing the external walls but do allow for connecting the buildings.
The site is located between a road (in the east), the seashore (in the west) and borders grounds on which there are similar resort homes belonging to the same firm (in the north) and an open field (in the south).
The buildings' location is marked on the drawing that appears on the next page.

Drawing for question 2-B.
It is permitted to answer on this page and then staple it to the test sheets. Additional drawings related to this question should be made on the large sheets.
The scale of the drawing is 1:100.



Section 3

In this section there are two questions of equal value. Answer only one of them. You must answer each part of the question you have chosen. Please read the whole question carefully before you begin. It is recommended that you organize the answer in a methodical, efficient and concise manner.
Time allotted for the answer: one and a half hours.

3-A. A poster: "Israel 2000 – Pilgrimage, scenery and recreation"

The year 2000 has been defined as a touristic-historic event and promises a large increase in the number of tourists-pilgrims to the sites in Israel which are sacred to Christians. The Ministry of Tourism is interested in distributing a poster called "Israel 2000 – Pilgrimage, Scenery and Recreation" emphasizing the following issues:

a. Pilgrimage.
b. Scenery, culture, heritage (Old Testament and New Testament scenery, historical sites).
c. Recreation and related activities.
d. Reconciliation, fraternity and bridging religions and nations.

1.1 Propose a poster that expresses three of the issues stated above. It is possible to present it as a combined poster or as a series of three posters which together form one poster.

1.2 In a concise manner explain the goals you wish to achieve by means of the poster (on the other side of the sheet).

3-B. A milk carton

Milk is a product consumed daily by different groups throughout the population. A milk carton is meant to keep the milk fresh and to accommodate transportation and storage. The carton is also meant to be eye-catching and appealing to the consumers to whom the product is marketed.

1.1 Propose three different milk cartons (half a liter or one liter in size), directed to different groups of consumers: a. babies; b. teenagers/students; c. adults. Choose the right size, material, shape and graphic design suited for each carton.

2.2 Write a brief explanation of your considerations in choosing your designs (on the other side of the sheet).

Appendix 2

Criteria for Evaluating the Architecture Selection Test (April 2000)

These directions were written for a specific test version, which is not the one described in appendix 1. They are used and presented here as another (third) source for the job/personlity analysis performed in chapter 2.

1. Demonstrating understanding of the question and providing a complete answer, that refers to the question and to all parts of it.

2. For answers requiring [the use of] visual skills:
This is the main component in evaluating each answer.
The higher the degree to which the following qualities are exhibited, the higher the score the answer receives:
   - 2.1 Perception of figures and their manipulation:
   Different cuttings of the cube and the cylinder. Composition of cubes and cylinders and/or their parts. Description of the road as a space including "floor", "walls" (and perhaps a "ceiling").
   Disregarding sizes (for example 80 centimeters - side of a cube), proportional distortion, disregard of matter, disregard of connections and connectors, disregard of proper use – all these will lower the score received.
   - 2.2 Visual memory:
   Will be expressed mainly in the answer to the question "the road as a meaningful place." Components and details in the road such as paving and vegetation will be evaluated. Components from the type detailed in paragraph 3 of the question.
   - 2.3 Technical understanding:
   Will be expressed mainly in the answer to the question "designing furniture/playing facilities…". The written explanation should also be considered.
   In the design of the poster the degree to which the message of the proposal is transmitted and the degree to which the page functioning as a poster (compared to a drawing in a book, for example) was understood, will also be evaluated.
   - 2.4 Clarity of presentation:
   The object and its components must be identifiable figuratively, as opposed to an "expressive" expression, like a long gray stain representing an asphalt road.
   - 2.5 Completeness and breadth of presentation:
   The objects and/or the space must be fully presented including structure and components. For example, description of the furniture only, without a description of the nursery in which they are posted, will decrease the score the answer receives.
   Presentations from different angles or separate descriptions of components which are not seen in the main drawing contribute to the mark.
   - 2.6 A graphic or written description of the use of the object or space contributes to the score.

3. For answers that require [the use of] verbal skills:
   3.1 The questions were presented in a manner that allows people, regardless of their native tongue, to give a correct written answer. That is why this component contributes only marginally to the score. However, one should regard a few aspects. Is the answer verbally correct? (For example: did the examinee state where the road is located? What are the factors affecting its character?) Do the words add information to the visual expression?
   3.2 The questions included in section 2 allow attentiveness to the physical, social and cultural environment to be expressed in words. A written answer which displays consideration of the environment will get a higher mark than a written answer which does not display this.
   3.3 Conciseness is required. An overly verbose answer lowers the mark received.

4. Appropriateness:
   Because there are many potential answers, the most appropriate answer will get a higher score. For example: furniture for children – a chair that suits a child's height is a more suitable answer than a TV table (which could also be placed in a nursery). A chair that suits a child's height is a more appropriate answer than a chair that suits an adult's height.

5. Complexity:
   Because there are many potential answers, a more complex answer will get a higher score. For example, in the subject of "the personality/the invention of the 20th century" a possible answer is the presentation of the invention of the birth control pills. An answer that offers the transition of women from their task as mothers and homemakers to professionals with careers would be a poorer example of an answer than one that would also present the general effect on the family, on women's sexuality, etc..

6. Originality and creative imagination:
   Because there are many potential answers, a more original answer will get a higher score. For example, a playing facility which is very similar (in form, in function) to that found in playgrounds will get a lower score than a playing facility which reveals innovation.

7. Efficient and systematic organization of the page:
   An efficient and systematic partition of the page, a balanced organization of the different components of the answer and use of page organization to contribute to the answer will receive a higher score. For example, placing the answers in a corner of the page with a reduction of the drawings' size will lower the score.
   Comment: This component has a large weight in section 2 of the test.

8. Diligence:
   Completion of all tasks and efficient organization of time.

Comment: It has been decided that unique works which call for unique criteria be discussed separately by a number of evaluators.

Appendix 3

Calculating criteria reliability

Each student's semesterial scores were randomly "split halved." Each half was averaged, and then a Pearson correlation coefficient between the two avereges was computed. Finally, a correction for length was performed according to the Spearman-Brown formula:

$$r_{tt} = \frac{2r_{hh}}{1 + r_{hh}}$$

Where:

r_{hh} – correlation between test halves.
r_{tt} – reliability of the complete test.

This calculation was performed for the scores of both theoretical Courses and Projects.

The estimation of the reliability of the combined average score was derived from the reliabilities of the project score and of the theoretical Course score, which are the components of the combined average score. This was done using the Feldt & Brennan (1989, p.117) formula:

$$r_{zz} = 1 - \frac{2^2\sigma^2_x(1-r_{xx}) + 1^2\sigma^2_y(1-r_{yy})}{3^2\sigma^2_z}$$

Where:

r_{zz} – reliability of the Combined average score.
r_{xx} – reliability of the average theoretical Course score.
r_{yy} – reliability of the average Project score.
σ^2_z – variance of the Combined average score.
σ^2_x – variance of the average theoretical Course score.
σ^2_y – variance of the average Project score.

Appendix 4

Calculating a regular composite score[17]

The composite score for all departments other than Architecture and Landscape Architecture was computed as follows:

1. The optimal average[18] B is computed.
2. The Preparatory Program average W, if there is one, is computed.
3. The higher of B and W is taken as the average D. That is, D=MAX (B, W).
4. The Comp. Score, S, is computed according to the formula below, where P stands for PET:

$$S= 0.611*D + 0.0616*P - 21.$$

Appendix 5

Correction for restriction of range[19]:

Notation:

Y = the criterion score.
X = the predictor the predictive ability of which we wish to estimate.
U = the variable according to which the selection was made (the composite score).

The variable U underwent direct selection and the variables X and Y underwent indirect selection, since they are correlated with U.
S and s designate standard deviations; R and r designate the correlation. Capital letters refer to estimations of parameters in the population (before selection) and lower case letters refer to statistics in the sample (after selection). Data relating to the population (S) are known only for the variable that underwent direct selection.

If the following assumptions are met:
1. the regressions of Y on U and of X on U are linear.
2. X's and Y's conditioned variances (and covariances) are independent of U (homoskedasticity).

then:

$$R_{xy} = \frac{r_{xy} + w_u r_{ux} r_{uy}}{\sqrt{(1 + w_u r_{ux}^2)(1 + w_u r_{uy}^2)}}$$

where:

$$w_u = (\frac{S_u^2}{s_u^2} - 1)$$

[17] From the 2001 Technion Bulletin: Information for the Applicats
[18] *Bagrut* average and additional scores.
[19] Adopted from Kennet-Cohen, Bronner & Oren, 1999a, after Linn, 1983.

Appendix 6

Raw Pearson correlations between predictors and first year scores (n=547)

	Predictors							
	AST	Bagrut	PET	E	V	Q	General Comp.	Arch. Comp.
Course Average	0.01	0.23	0.23	0.16	0.08	0.24	0.31	0.30
Project Average	0.15	-.02	0.05	0.03	-.00	0.07	0.01	0.14
Combined Average	0.09	0.16	0.19	0.13	0.06	0.22	0.23	0.29

Appendix 7

Partial coefficients (β) and multiple correlations (R) in multiple regression for predicting first FY scores, based on raw correlation coefficients

Predictors	Three-variable model				Five-variable model					
	AST	Bagrut	PET	R	AST	Bagrut	E	V	Q	R
FY Criteria	β				β					
Course	0.14	0.25	0.25	0.39	0.14	0.23	0.15	0.04	0.22	0.44
Project	0.18	0.02	0.09	0.27	0.18	0.01	0.06	-0.01	0.09	0.31
Combined	0.19	0.20	0.24	0.35	0.20	0.18	0.14	0.03	0.22	0.41