דוח המרכז

268

A Rationale for the Design of the Psychometric Entrance Test

Yigal Attali
Chanan Goldschmidt

December 1999

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 268
ISBN:965-502-072-X

21 ISBN 072

A Rationale for the Design of the Psychometric Entrance Test

Yigal Attali
Chanan Goldschmidt

December 1999

We thank Ruth Fortus, Naomi Gafni, Yoav Cohen and Jerry Levinson for their time and helpful comments.

A Rationale for the Design of the Psychometric Entrance Test

Preface

The Psychometric Entrance Test (PET) is produced by the National Institute for Testing and Evaluation (NITE) and used in the selection of candidates for higher education in Israel. PET consists of about 220 short multiple-choice questions and the duration of the test is three hours and twenty minutes.

PET is composed of three subtests: Verbal Reasoning, Quantitative Reasoning, and English. The Verbal Reasoning subtest includes questions in the domain of reading comprehension, vocabulary, and analytic reasoning. The Quantitative Reasoning subtest includes questions based mainly on the Mathematics curriculum taught in elementary and high-school. The English subtest includes questions aimed at assessing the ability to read and understand English texts at an academic level.

Many questions arise from this short description of PET. For example, the same PET score is used for the selection of candidates for Biology studies, Philosophy, and Law. Surely the abilities that are needed to succeed in these different fields of study are different, so why is the same test being used?

A similar point could be made about the Quantitative Reasoning section. Most students do not need mathematical knowledge for their studies, so it is odd that this domain is assessed for all candidates.

There are other abilities that are relevant for many academic fields and are not assessed in PET, such as writing skills, motivation, and creativity, to name a few.

These and other questions concern consumers of PET, both university candidates and decision-makers in the universities. This paper will try to address these issues by explaining why the current test content was chosen, why other contents do not appear in the test, and why the format of the test is as it is.

Introduction

This paper presents a rationale for the design of PET. The design of a test, however, is only a part of a larger picture. Any assessment process involves a number of interconnected stages, and Willingham and Cole (1997) identify four of these: design, development, administration, and use. Each stage is related to the other, with some connections stronger than others. Design and use are closely related, as are development and administration. Test design is intimately connected with test use, because the choice and nature of the constructs measured in a given test depend on the intended use of the test and the possible effects of its use.

The most important choices in test design concern the knowledge and skills to be measured and how to measure them. In order to do that, it is first necessary to establish the purpose of the test, which constructs will be useful in serving that purpose, and which assessment format to use. Next come more detailed decisions regarding the particular types of items.

The main purpose of PET is to function as an admissions test for the Israeli universities. PET is used to help admissions officers select those students who have the highest probability of succeeding in their studies. Therefore, the question is what psychological constructs, or aptitudes, are relevant for learning, development, and achievement in the university. Recent conceptions of aptitude focus on the initial states of persons that influence later developments (Snow, 1992). The notion of initial states covers a broad array of personal characteristics, including cognitive and affective characteristics (Snow, 1991).

Snow (1989) presented a sketch of the psychological constructs involved in learning and development. He identified five categories: *conceptual structures of declarative knowledge*; *procedural skills* involved in learning, thinking, and reasoning; *learning strategies*, styles, and tactics; *self-regulatory functions*; and *motivational orientations*.

In order to establish which of these constructs will be most useful in serving the purpose of PET, it is important to specify the context in which PET's scores are used. PET scores are only one part of the final admissions score. The high-school matriculation scores (the Bagrut) of the applicant are also taken into account. The Bagrut scores are obtained by assessing achievement in specific domains, such as Mathematics, English, History, Social Studies, Physics, and many others. Thus, the Bagrut's main focus is the first aptitude identified by Snow (1989), namely, the *conceptual structures of declarative knowledge*. In addition, the Bagrut indirectly measures aspects of the last two conative aptitudes (i.e., conscious willingness to act or exert effort): self-regulation and motivation. Whether or not students have applied themselves to their studies consistently over a long period of time will most likely have an effect on their scores.

Historically, the Bagrut scores were the only scores that were taken into account for admissions purposes. PET was created for two reasons: the first is the difficulty in comparing different Bagrut scores, both because each student may take tests in different subjects, and because the tests are not comparable from year to year. Furthermore, the final score on each Bagrut test is composed of a score on an external exam, administered by the Ministry of Education, and an internal score, administered by the teacher. The second reason is the desire to give students a "second chance" for acceptance to higher education.

Thus, it can be said that the Bagrut constitutes the basis of the final admissions score, and that PET is a supplementary score. The intended use of PET as an admissions test and the existence of Bagrut scores are the primary determinants of PET's design. Thus, PET should aim at assessing the aptitudes that the Bagrut does not focus on: the *procedural skills involved in learning, thinking, and reasoning*, or, in other words, the previously developed abilities and skills that can be capitalized upon as *general tools* for learning in new domains.

Conditions and Limitations Affecting the Nature of PET

This section presents a set of conditions and limitations which affect the nature of PET, and the specification of skills that PET is intended to measure.

Comparability of Scores

Comparability of scores is important for any test, and particularly important for PET. As mentioned above, the difficulty in comparing Bagrut scores is one of the reasons for PET's existence.

Comparability of scores is attained by standardizing the test and ensuring equivalency of the different test forms. A standardized test is one for which the conditions of administration and the scoring procedures are designed to be the same in all uses of the test (Millman & Greene, 1989). The conditions of administration include the physical setting, the instructions given to examinees, the test materials, and the time factor. Scoring procedures include both derivation and transformation of raw scores. Comparability of scores across time is also very important: scores that are obtained on test forms taken months or years apart should be equivalent to one another.

There are a number of factors that have direct impact on the degree of comparability of scores *and* on test content, and therefore should be discussed here.

Test Length

It is well known that the length of a test has an effect on the test's reliability and validity. There is a high positive correlation between the number of items on a test and its reliability. Other considerations, however, constrain the number of items that can be used on a test. Large-scale assessment requires that the test be administered during the course of one day. The factor of fatigue must also be taken into account: requiring examinees to reason or write for extended periods of time will result in a decline in performance. Thus, the time allotted for the test should not exceed four or five consecutive hours.

The time limit has implications for the *number* of content areas that PET may assess. Because each content area tested consumes some testing time, the number of contents is limited. For example, spatial ability is involved in learning and reasoning, at least in some academic fields of study, but is not tested in PET, at least in part because of the time limit and the need to weigh the relative effectiveness of each skill for estimating academic success.

Item Format

One solution to the problem of trying to maximize the number of items without exceeding a particular time limit is to use items that can be answered in a relatively short amount of time, are clear and easy to understand and do not require long or complex instructions. One item format that satisfies these constraints is short multiple-choice (MC) items.

There is another reason why the MC item format is the one used in PET: the objective way in which scores are derived in MC tests make a substantive contribution to the standardization of the test.

Since PET is committed to a paper-and-pencil MC format, testing certain verbal skills is difficult or impossible. It is clear, for example, that testing oral skills directly is not a possibility. Testing writing skills posits similar problems. Directly assessing the student's ability to write well is not possible due to the format of the test, and although numerous types of discrete items have been developed to measure such writing skills as grammar, usage, sentence structure, and so on (these types are sometimes referred to as indirect measures), they are not included in PET.

In general, a MC test is restricted to *recognition* tasks as opposed to *production* tasks. Writing is a skill that is fundamentally productive. Other fundamentally productive skills include creative skills, which are also not tested in PET.

Other Considerations Affecting the Content of PET

English Proficiency

It is the policy of Israeli universities that students have at least a minimal knowledge of English before commencing their studies, and attain a certain level of comprehension in English before graduating. Students who have not yet achieved this level are given remedial English courses. One of the purposes which PET must fulfill is the classification of students according to their ability to read and understand English texts at an academic level, thus enabling their placement in the appropriate classes. For this reason PET includes as one of its subtests a test of English as a foreign language.

Breadth of Study Fields

In Israel, preparatory schools do not exist; students begin studying in the department of their choice (even in Law or Medicine) from their first year. There is no Israeli equivalent of the LSAT or GRE to select candidates for graduate school. All the universities (and most of the colleges and other institutions of higher education in Israel) use PET as a selection tool. This state of affairs implies that PET should measure abilities which are relevant to a wide range of curriculae.

Self-Selection of Students

In Israel, students apply to specific departments or schools in every university. Naturally, students prefer some departments over others. This self-selection of students for various departments, in addition to the different capacities of those departments, means that departments will vary greatly in their cutoff points for applicants. Consequently, PET must be able to assess both high-ability and low-ability candidates, and thus it must include items ranging from the very easy to the very difficult.

Implications of the Coaching Industry for Test Content

According to results from feedback questionnaires of PET examinees, about 80% of all examinees report that they participated in commercial preparation courses before taking the test. The abilities that PET is supposed to assess develop over many years through many and diverse learning experiences, both in and out of school. Therefore, practice and learning could contribute to performance on the test. However, performance should not be influenced by learning "tricks" and exposure to the specific content of the test.

Testing in Different Languages

One of the unique conditions under which PET must operate is the existence of a large number of students who are not native speakers of Hebrew. Most of them speak Arabic or Russian, and so equivalent forms of the tests must be created in these and other languages. Thus, the items that are used in PET must be translatable.

Unbiased Items

The items in PET should not be biased against various subgroups of examinees. Thus, the items that are used in PET should, as far as possible, provide examinees from different subgroups with a comparable opportunity to demonstrate the knowledge and skills they have acquired that are relevant to the purpose of the test.

General Skills Assessed in PET

The preceding section discussed some of the considerations that influence the type of skills assessed in PET. Consequently, PET assess two general skills involved in learning and thinking. This section will discuss several aspects of the two general skills involved in learning and thinking that are assessed in PET: reasoning skills and verbal skills.

Reasoning skills

Reasoning pertains to the process of drawing conclusions from principles and from evidence (Wason & Johnson-Laird, 1972), moving on from what is already known to infer a new conclusion or to evaluate a proposed conclusion. Reasoning is abundant in our life: we reason when we predict the future, evaluate the past, and think about the present; when we read or listen and when we write or talk; when we engage in activities as diverse as driving, cooking, and understanding the reason for an electrical blackout.

In order to show that assessing reasoning skills would contribute to the stated purpose of PET, it is necessary to demonstrate that reasoning skills are important for *learning*. This can be accomplished by describing how the *use* and *acquisition* of knowledge in many domains build upon the concept of reasoning. But the demonstration that reasoning skills are important for learning is not enough; an important task is to specify how, or by what means, these skills should be assessed. This is a crucial matter precisely because reasoning skills are *general*, and therefore content specification for this kind of test is particularly important.

Types of Reasoning

Reasoning is often divided into two types - deductive and inductive. In *deductive reasoning*, people are trying to determine what conclusion, if any, necessarily follows when certain given statements are assumed to be true. Deduction is used whenever we want to follow rules, whether general or universal. For example, deduction is used in planning electrical circuits and in cooking, because in order to operate successfully in both domains one needs knowledge of a complicated system of rules.

Inductive reasoning is the process of reasoning from specific facts or observations to reach a likely conclusion that may explain the facts. Induction encompasses all inferential processes that expand knowledge in the face of uncertainty, and it may be applied to situations requiring causal and categorical inferences, or, more generally, when we want to induce or impose *structure* on observations.

Reasoning and the Use of Knowledge in Problem Solving

One of the central uses of knowledge and skill is to solve problems. The ability of a person to solve problems in a specific knowledge domain is often the best indication of him or her being an expert in the domain (for example, think of technicians, lawyers, computer programmers, physicians, fire-fighters, chess players, diplomats, etc.).

Although problems in different domains have different characteristics, the solution of many problems requires the application of general reasoning skills. Greeno (1978) and Greeno and Simon (1988) classified problems into four types: (1) problems of transformation, in which an individual must have a clear definition of the present situation (problem) and the desired goal situation (the solution), and then must implement a sequence of operations to reach the goal; (2) problems of arrangement, in which an individual has an assortment of elements and a general idea of the goal and must arrange the elements in such a way as to reach the goal; (3) problems of inducing structure, which involve inductive reasoning; and (4) problems of evaluating deductive arguments, which involve deductive reasoning.

It is clear that all these problem types involve the application of inductive and deductive processes. For example, in problems of transformation the order and nature of the operations that are needed for solution are often determined by rules. Moreover, the tools for solving problems are deductive and inductive in nature: algorithms and heuristics (e.g., for dividing fractions) are deductive systems, and analogy (e.g., the analogy between electrical circuits and water flowing in pumps) is an instance of induction. In conclusion, it can be said that in using our knowledge we often apply reasoning skills.

Knowledge Acquisition and Organization of Knowledge

How are reasoning skills involved in knowledge *acquisition*? The answer to this question is related to the differences between experts and novices in a particular knowledge domain. How does knowledge, particularly expert knowledge, enhance problem solving? Research on expert knowledge (Chase & Simon, 1973; de Groot, 1965) suggests that what differentiates the experts from the novices in a particular domain is the *organization* of knowledge. Not only do experts have more knowledge, but their knowledge is better organized, thus providing them with a greater capacity to construct useful *representations* of the problem at hand. Problem representation is a significant part of the solution process: how (and if) a problem is solved depends fundamentally on how it is represented. The principle difficulty in solving insight problems, for example, is that they naturally stimulate wrong or non-useful representations.

In other words, experts seem to have tied (declarative) knowledge to the conditions of its use - they seem to have *proceduralized* their knowledge. This distinction between declarative and procedural knowledge represents the difference between knowing that something is the case and being able to do something, often automatically and effortlessly. In the field of mathematics, for example, George Polya claimed that mathematics consists of information and know-how. Regardless of the amount of information the students are exposed to, they have to know how to use it. "To know mathematics is to be able to do mathematics" (Polya, 1969/1984, p. 574). "What is know-how in mathematics? The ability to solve problems" (Polya, 1981, p. xi).

Reasoning and the Acquisition of Knowledge

How do students proceduralize their knowledge? How is this know-how acquired? There is agreement in the problem solving literature that proceduralization of knowledge is accomplished through experiences with many *examples* (Anderson, 1993; Simon & Zhu, 1988; VanLehn, 1986, 1990). That is, by following the steps in a worked-out example, students may generalize or abstract the correct procedure for the given skill. This means that principles, rules, algorithms, and other deductive structures are learned and internalized through an inductive process.

A specific kind of inductive reasoning - analogical reasoning - is particularly important in this context. Analogical reasoning can be characterized as a process of finding an appropriate source analog for the target problem at hand, and forming a mapping between the source and target problems. Many findings (e.g., Anderson & Thompson, 1989; Novick & Holyoak, 1991) point to the importance of forming analogies for learning, the end result of which is the induction of a set of more abstract rules that embodies the relationship between the source and target problems.

Well- and Ill-Structured Knowledge Domains

So far the discussion on the use and acquisition of knowledge has been general, and no distinction was made between domains as different as history and physics. In fact, there are important differences between problem solving in physics and problem solving in history, differences that are relevant for the assessment of general reasoning skills.

On the one hand, the domains of natural sciences, mathematics, engineering, and computer sciences all have in common the fact that operating in these domains is accomplished by the use of *formal languages*, like programming languages, the mathematical language, and logical languages. These languages are created for a specific use and are characterized by a perfect syntax (rules that determine what are admissable expressions in the language) and perfect semantics (rules that determine the meaning of expressions). It can even be said that learning in these fields is expressed as a greater mastery of these languages. Consequently, problems in these domains are usually well-structured: the premises of the problem are clear, there are established procedures for solving the problem, and there is one solution for the problem.

On the other hand, operating in the fields of social sciences, medicine, law, management, and the humanities is not possible without the use of natural language, with all its ambiguity, inconsistency, and multiple meanings (polysemy) in production and understanding. Trying to express ideas in these fields in some formal language is fruitless.

Consequently, problem solving in these domains resembles in some ways *informal* or *everyday reasoning* (Galotti, 1989): the intellectual activities that compose the thinking done in our everyday lives, such as planning, evaluating arguments, and choosing options. In these problems (also called ill-structured problems), some premises of the problem are implicit, and some are not supplied at all. There are typically several possible solutions to the problem that vary in quality, and there are no established procedures for solving the problem. These domains are knowledge-rich: the problem solver has to bring a large body of *background knowledge* to bear in solving the problem. Thus, in these problems knowledge representation and determining just what information is relevant is often a crucial determinant of successful solution.

Assessment of Reasoning Skills

The previous discussion intended to show the relevance of reasoning skills to the use and acquisition of knowledge, thereby explaining why assessing these skills to obtain an indication of the learning ability of the university applicant is important. However, a problem arises: reasoning skills are expressed when actually solving problems in some specific knowledge domain. But in order to solve problems in a domain an examinee must have experience in that domain - or at least the opportunity to study the domain. So, to assess the *general* reasoning skills of examinees, an evaluation of problem solving in some *specific* knowledge domain must be made.

There are two ways to overcome this problem. The first is to assess the problem solving abilities and understanding in a knowledge domain with the following two properties: it must have a complex rule space, and all applicants must have a great deal of experience in the domain, both in learning and in problem solving. Such an assessment has the status of an *indicant*: if applicants, who have had years of opportunity to learn and develop their problem solving skills in this domain, are able to successfully solve problems in the domain, it is an indication that they have highly-developed reasoning skills that could allow them to learn and become an expert problem solver in *other* complex domains, for example in the university.

There is one knowledge domain that has the above two properties: the mathematical domain. It is complex and is learned intensely during long school years, starting from first grade. It is also a prime example of a structured field, and thus it should serve as a particularly good indicant for the learning abilities of other structured domains. For these domains, mathematical problem solving skills are more than an indirect indication of the existence of reasoning skills - they are evidence of mastery of the mathematical language, which is a basic formal language. Many structured domains use this language, and many other formal languages are based on it. Learning physics, for example, is not possible without some understanding of mathematics.

The second way in which PET can use the prior experience of applicants to assess their reasoning skills is based on the resemblance mentioned above between problem solving in ill-structured domains and informal or everyday reasoning. All applicants have long years of experience in everyday reasoning, and consequently have had the opportunity to develop two skills that are particularly relevant to problem solving in these domains, and are in fact instances of inductive thinking: argumentation and explanation skills.

Argumentation skills are the ability to understand, analyze, and evaluate arguments: knowing what kind of evidence will support or refute a hypothesis; recognizing central arguments in a thesis; identifying both stated and unstated assumptions in an argument; and recognizing fallacies and contradictions in arguments.

Explanation skills are the ability to construct explanations, assess the soundness and consistency of inferences and conclusions, and generate alternative explanations and counterexamples to arguments.

The assessment of argumentation and explanation skills provides a different kind of information about the applicant than does the assessment of mathematical problem solving. It serves less as an indication of learning ability of new domains and more as an indication of mastery of the basic means by which students would understand and learn in ill-structured domains. Consequently, the tasks presented to the examinees during the test would often be similar to actual tasks that students would encounter when studying such fields as psychology and history, unlike the mathematical problems that the applicant will have to solve during the test, which he or she will seldom encounter in academic studies (even in studies of mathematics). The examinees would have to explain a puzzling phenomenon, criticize an argument, and so on.

Verbal skills

Verbal skills represent mainly what one may call competence or knowledge in the use of language. Carroll (1993) describes four classes of language skills that are distinguished on the basis of receptive versus productive skills in both oral and written language. Oral language includes listening as a receptive skill and speaking as a productive skill. Written language includes reading as a receptive skill and writing as a productive skill.

Verbal skills are indispensable for learning because any kind of *comprehension* of any mode of learning material is mediated by these skills - particularly listening and reading. On the other hand, indication of learning usually involves some kind of language production. PET items assess only reading skills (the reason for this was explained above), therefore a description of comprehension processes in reading follows.

Comprehension in Reading

Although verbal *comprehension* includes both reading and listening, cognitive models of oral comprehension tend to resemble models of reading comprehension. Comprehension processes are the means by which we make sense of what we read. What are the main comprehension processes? The first step in comprehension is to understand the meaning of the words. *Semantic encoding* is the process by which we translate sensory information into a meaningful representation that we perceive, based on our understanding of the meaning of the words. Individual word meanings would then have to be combined into fundamental ideas (Kintsch, 1990; Kintsch & Van Dijk, 1978). But before true understanding of the text can occur these ideas must be tied to the reader's prior knowledge to form a mental model of the text that simulates the world that the individual words and phrases describe (Kintsch, 1986; Perfetti, 1985, 1986). Understanding requires a constant coordination and updating of this model. The resemblance of this description to reasoning is not accidental. The same processes must be present in comprehending an argument or an explanation. This suggests that these comprehension processes are related to the reasoning processes used to solve ill-structured, semantically-rich problems. Consequently, when assessing what may be the most natural task in this context, namely the reading comprehension of academic texts, it is difficult, even conceptually, to separate reading skills from reasoning skills.

Importance of Vocabulary for Comprehension in Reading

The processes of semantic encoding of words and the higher-order processes of forming and updating models of the text interact in the continuous process of reading from the text, for example, through the allocation of processing resources among them. Efficiency in all components makes for high reading ability. Inefficiency in one taxes attention, reducing the resources available for the other component processes. Moreover, at the academic level, the evidence suggests that the poor reader in the university is likely to show deficits in all verbal abilities, including syntax, vocabulary, spelling, and composition (Snow & Lohman, 1989).

For example, it is clear that vocabulary, or knowledge of word meanings, is very closely tied to the ability to comprehend a text: readers simply cannot understand a text well unless they know the meanings of the component words. But another way in which having a large vocabulary contributes to text comprehension is through learning from context. If we cannot semantically encode a word because its meaning does not already exist in memory, we must find another way in which to derive its meaning, for example, by noting the context in which it appears. Werner and Kaplan (1952) proposed that people learn most of their vocabulary indirectly, not by using external resources such as dictionaries or teachers, but by figuring out the meanings of words from the surrounding information.

Daalen-Kapteijns and Elshout-Mohr (1981) proposed that subjects first generate a schema or hypothesis for the meaning of an unfamiliar word. Based on an active search in the text, the schema is adjusted or reformulated in order to confirm it. In their research, low-verbal subjects were less likely than high-verbal subjects to use this strategy of schema-guided search, possibly because they were not able to activate the appropriate knowledge schema for the words.

These findings suggest that processes of word acquisition are closely related to processes occurring during normal reading and comprehension of a text.

Analysis of PET Item Types

In this section the relevance of PET item types to verbal and reasoning abilities will be outlined.

Verbal Reasoning Section

The overlap between reasoning and verbal abilities is manifested in the verbal item types appearing in PET. Some item types are concerned primarily with the ability to analyze and understand complex written material, other item types mainly involve the ability to think clearly and systematically, and still other item types reflect a mixture of such abilities.

The current PET verbal section consists of six item types: words and expressions, analogies, sentence-completions, letter-exchange items, analytical items, and reading comprehension.

Words and Expressions
The purpose of this item type is to directly assess the vocabulary of examinees. The items appear in a number of forms: items dealing directly with the meaning of words or expressions; sentence-completion items with one blank; antonyms; and items in which examinees have to choose one option whose meaning is distinct from that of the other options.

Recently, efforts have been made to include more of the sentence-completion type of item. This more accurately reflects a focus on testing the comprehension processes in reading by measuring vocabulary knowledge in context rather than by merely asking for a definition of a word. Sentence-completion items are also less susceptible to memorization of word lists and call for more ecological preparation methods of constructing meaning from text.

Analogies
As has been noted previously, analogical reasoning is a powerful tool in learning and understanding. By using analogies we transfer (map) abstract relations from familiar base domains of knowledge to less familiar target domains. These abstract relations are of many kinds, but the verbal analogies appearing in PET focus on a specific kind, namely the semantic meaning of concepts. These analogies test the ability to define the relationship between two concepts and to recognize a similar relationship in other pairs of concepts.

Sentence-Completions
Sentence-completion items consist of a sentence with three or four blanks. A word or words are missing from each of the blanks. These items emphasize understanding of the logical and semantic relationships within a complex sentence. The connections between different parts of the sentence are of various types: one part could explicate another part, exemplify it, negate it, etc. In decoding the nature of the relation between the parts of the sentence, one must pay particular attention to the prepositions in the sentence, because their meaning establishes the type of connection. After filling in the blanks correctly, the entire sentence should constitute a coherent argument. Thus, the ability to analyze and understand arguments is needed for solving these items.

Letter-Exchange Items
These items are relatively new in PET (they were introduced in 1993). They were developed at NITE and are based upon a morphological feature of Semitic languages not shared by Indo-European ones, namely, the fact that most of the vocabulary in Hebrew - all verbs and most nouns and adjectives - can be characterized as a combination of Root + Pattern. The root is most typically composed of three

consonants, and it carries the semantic core of the words formed by it; the patterns take the form of vocalic and syllabic additions to the root, and they serve to modify the core meaning of the root.

The letter-exchange items are composed of four sentences. In each sentence one word is altered by changing its root letters into a standard template (the letters **p.t.l.**). In three of the four sentences the standard template stands for the same three letters. In the remaining sentence the template replaces another root. The examinees have to identify this sentence.

The process of solving these items is composed of two operations that are executed iteratively: generating hypotheses about the root that can replace **p.t.l.** in a given sentence, and checking these hypotheses by using that root in the other sentences.

Generating hypotheses about the possible roots that would form a plausible sentence is mainly determined by vocabulary. The ability to be flexible in one's thinking, by rejecting unsuitable hypotheses and generating new ones, is also tested in this item type.

Analytical Items

The analytical items assess the ability to analyze information and then make inferences from it regarding a particular conclusion. There are two kinds of conclusions in these items: conclusions that are *necessarily* true (or false), given the information, and conclusions that are *reasonable* or *possibly* true, given the information. The first kind appears in items that are deductive in nature, and the second kind appears in items that are more concerned with inductive and informal reasoning.

There are two main types of deductive problems and both are present in PET. In one type, called categorical syllogisms, people are required to determine what conclusion, if any, must follow from certain assumptions about category membership. The other type, called propositional problems, concerns the evaluation of the truth of arguments consisting of sequences of simple statements linked by connectives such as *and*, *or*, *not* and *if* . . . *then* to form compound statements.

In induction, two categories of task can be distinguished. These may be labeled *hypotheses testing* and *hypotheses generation*. In hypotheses testing, people are required to determine the implications, if any, of certain observations with regard to the truth of possible generalizations (hypotheses). In hypotheses generation, the person has or can obtain information on the objects of interest and seeks to make a plausible generalization. The format of PET items constrain the items to hypothesis testing: the examinees are required to perform a variety of tasks, such as judging the plausibility of conclusions, recognizing assumptions with respect to certain conclusions, or analyzing the effects of additional information on a conclusion.

Reading Comprehension

This item type reflects the conception of the skilled reader as one who constructs meaning from a text, as opposed to simply decoding what is on the page. This view recognizes the active role of the reader as one who brings to the task prior knowledge on the topics about which he or she is reading and uses all available information to determine the writer's intent.

The items reflect the process that a reader goes through while deriving meaning from the text. They assess the test taker's ability to interpret, synthesize, analyze, and evaluate the reading material, and thus measure higher order analytical and evaluative skills.

The passages are intended to be well written and readable, engaging, not overly dense or technical, and with a sufficient amount of contextual information.

Some passages include two points of view which oppose, support, or in some other way complement one another. Some of the items on these passages will assess the examinee's ability to compare or contrast the two points of view, use information from one to interpret information in the other, and identify assumptions they share or pivotal differences between them.

The content of the reading selections is scrupulously balanced and is drawn from the humanities, natural and physical sciences, and social sciences. The balance of passage content is important, since not only do examinees have background knowledge in a variety of subjects, but they will also vary in their future fields of study.

Quantitative Reasoning Section

The rationale for including a quantitative reasoning section in PET is to assess the reasoning skills of students through their problem solving ability in a central domain that is well-structured and demands extensive reasoning in a complex rule space. Such an assessment can be diagnostic of the examinee's learning potential in many academic domains. This is because the processes through which mathematical knowledge is proceduralized and problem solving is enhanced are similar to processes which occur in other domains, especially the physical sciences (Snow & Lohman, 1989).

Mathematical problem solving involves the understanding of mathematics in ways that enable students to reason meaningfully with and about mathematical concepts and principles. Mayer's (1982, 1985) theory of mathematical problem solving may be used to demonstrate the appropriateness of PET quantitative items to current views on the assessment of problem solving skills.

Mayer hypothesized that problem representation is composed of two parts: problem translation (understanding each statement in the problem) and problem integration (assembling a coherent representation from the story-problem propositions). Translation requires linguistic knowledge (how to parse sentences, what various words mean) and factual knowledge (general world knowledge such as units of measurements, time and currency related measurements), whereas integration requires schema knowledge (knowledge about problem types such as distance-speed-time problems).

Mayer (1982) divided problem solution into two phases: solution planning and solution execution. Solution planning relies on general strategic knowledge and heuristics, but depends on how the problem was represented. Finally, solution execution depends on how well the student has correctly automated algorithms for solving equations, multiplying numbers, and the like. This is the most obvious, yet, in a sense, the least important step in the process.

There are three types of items in the Quantitative Reasoning section: questions and problems, quantitative comparisons, and graph or table comprehension. These items focus on problem representation and planning skills, and emphasize mathematical understanding and problem solving, rather than computation, in the domains of arithmetic, algebra, geometry, and comprehension of graphs and tables. This is accomplished by developing items that are non-routine problems, and by making use of diagnostic distractors which reflect errors in the selection of algorithms and in the understanding of the problem, rather than in the execution of algorithms. Assessment of understanding rather than computation is further emphasized by providing a list of mathematical definitions and formulas in each section.

English Section

This section consists of three item types that test command of the English language in terms of the ability to read and understand English texts at an academic level. Sentence-completion items assess the English vocabulary of the examinee and the ability to use words in the context of a sentence. In Restatement items, a sentence is presented, followed by four possible restatements of that sentence, only one of which is correct. These items are intended to test vocabulary, syntax, and the ability to understand the relationships between different parts of a sentence. Reading comprehension items assess the ability to understand short passages. The items related to a passage could touch upon a word, a sentence, or a larger part of the passage.

Summary

This paper examined the rationale for the design of PET, the aptitudes it measures and the way it measures them. As a consequence of the use of the Bagrut as part of the admissions score, PET focuses on the assessment of reasoning and verbal skills, both general skills that are relevant for learning and achievement in academic studies. This is accomplished by generating a standardized test composed of MC items in three subtests: Verbal Reasoning, Quantitative Reasoning, and English. The discussion of the item types included in these subtests points to their suitability as indicants of reasoning and verbal skills.

The results of many validity studies show that PET is a good predictor of success at university, and that the combination of the Bagrut and PET as a predictor is better than using either of them separately.

Nonetheless, PET does not claim to assess all the general skills relevant for learning and achievement. The development of new tools for the assessment of general skills, such as a writing test, should continue to be a focus of research at the National Institute for Testing and Evaluation.

References

Anderson, J.R. (1993). Problem solving and learning. *American Psychologist, 48*, 35-44.

Anderson, J.R., & Thompson, R. (1989). Use of analogy in a production system architecture. In S. Vosniadou and A. Ortony (Eds.), *Similarity and analogical reasoning* (pp. 267-297). Cambridge, MA: Cambridge University Press.

Carroll, J.B. (1993). *Human cognitive abilities*. Cambridge: Cambridge University Press.

Chase, W.G., & Simon, H.A. (1973). The mind's eye in chess. In W.G. Chase (Ed.), *Visual information processing* (pp. 215-281). New York: Academic Press.

Daalen-Kapteijns, M.M., & Elshout-Mohr, M. (1981). The acquisition of word meanings as a cognitive learning process. *Journal of Verbal Learning and Verbal Behavior, 20*, 386-399.

de Groot, A.D. (1965). *Thought and choice in chess*. The Hague, Netherlands: Mouton.

Galotti, K.M. (1989). Approaches to studying formal and everyday reasoning. *Psychological Bulletin, 105*, 331-351.

Greeno, J.G. (1978). A study of problem solving. In R. Glaser (Ed.), *Advances in instructional psychology: Vol 1* (pp. 13-75). Hillsdale, NJ: Lawrence Erlbaum Associates.

Greeno, J.G., & Simon, H.A. (1988). Problem solving and reasoning. In R.C. Atkinson, R. Herrnstein, G. Lindzey, R.D. Luce (Eds.), *Stevens' handbook of experimental psychology* (rev. ed.), New York: John Wiley & Sons.

Kintsch, W. (1986). Learning from text. *Cognition and Instruction, 3*, 87-108.

Kintsch, W. (1990). The representation of knowledge and the use of knowledge in discourse comprehension. In C. Graumann & R. Dietrich (Eds.), *Language in the social context.* Amsterdam: Elsevier.

Kintsch, W., & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*(5), 363-394.

Mayer, R.E. (1982). Memory for algebra story problems. *Journal of Educational Psychology, 74*, 199-216.

Mayer, R.E. (1985). Mathematical ability. In R.J. Sternberg (Ed.), *Human Abilities: An information-processing approach* (pp. 127-150). San Francisco: W.H. Freeman.

Millman, J., & Greene, J. (1989). The specification and development of tests of achievement and ability. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 335-366). New York: Macmillan.

Novick, L.R., & Holyoak, K.J. (1991). Mathematical problem solving by analogy. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 398-415.

Perfetti, C.A. (1985). Reading ability. In R.J. Sternberg (Ed.), *Human Abilities: An information-processing approach* (pp. 59-81). San Francisco: W.H. Freeman.

Perfetti, C.A. (1986). *Reading ability.* New York: Oxford University.

Polya, G. (1981). *Mathematical discovery* (Combined ed.). New York: Wiley.

Polya, G. (1984). Fundamental ideas and objectives of mathematical education. In G.C. Rota (Ed.), *George Polya: Collected papers: Vol. 4. Probability; combinatorics; teaching and learning in mathematics* (pp. 569-578). Cambridge, MA: MIT Press. (Reprinted from *Mathematics in Commonwealth Schools*, 1969, pp. 27-34)

Simon, H.A., & Zhu, X. (1988). Learning mathematics from examples and by doing. *Cognition and Instruction, 4*, 137-166.

Snow, R.E. (1989). Toward assessment of cognitive and conative structures in learning. *Educational Researcher, 18*, 8-14.

Snow, R.E. (1991). The concept of aptitude. In R.E. Snow & D.F. Wiley (Eds.), *Improving inquiry in social science* (pp. 249-280), Hillsdale, NJ: Erlbaum.

Snow, R.E. (1992). Aptitude theory: Yesterday, today, and tomorrow. *Educational Psychologist, 27*, 5-32.

Snow, R.E., & Lohman, D.F. (1989). Implications of cognitive psychology for educational measurement. In R.L. Linn (Ed.), *Educational measurement* (3rd ed., pp. 263-331). New York: Macmillan.

VanLehn, K. (1986). Arithmetic procedures are induced from examples. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 133-179). Hillsdale, NJ: Lawrence Erlbaum Associates.

VanLehn, K. (1990). *Mind bugs: The origins of procedural misconceptions*. Cambridge, MA: MIT Press.

Wason, P.C., & Johnson-Laird, P.N. (1972). *Psychology of reasoning: Structure and content*. London: B.T. Batsford.

Werner, H., & Kaplan, E. (1952). The acquisition of word meanings: A developmental study. *Monographs of the Society for Research in Child Development* (No. 51).

Willingham, W.W., & Cole, N.S. (1997). Gender and fair assessment. Mahwah, NJ: Lawrence Erlbaum Associates.