DEVELOPMENT OF THE SYNDICATE'S

HIGHER EDUCATION

APTITUDE TESTS

by

ALEC FISHER

April 1992
Development of the Syndicate's Higher Education Aptitude Tests.

CONTENTS

1. Background.
   1.1. The Law Studies Test.
   1.2. The Academic Aptitude Profile (AAP) and the Higher Education Aptitude Test (HEAT).
   1.3. The Assumptions Behind Aptitude Testing.
   1.4. How the AAP and HEAT differ from North American tests.

2. Logical Reasoning.
   2.1. Origins.
   2.2. How Logical Reasoning is Conceived.
   2.21. Question types.
   2.3. Comment.
   2.31. Face Validity.
   2.32. Judging Reliability of Evidence.
   2.33. 'Productive' Logical Reasoning Capacities.
   2.34. Dispositions.

   3.1. Origin and Purpose.
   3.2. How Mathematical Reasoning is Conceived.
   3.21. Question types.
   3.3. Comment.
   3.31. Testing Aptitude or Knowledge?
   3.32. The Assumptions Behind Aptitude Testing.
   3.33. 'Productive' versus 'Schema-driven' Thinking.
   3.34. General Conclusions.


5. Reading Comprehension and Communicative Skills.
   5.1. Origins.
   5.2. How Reading Comprehension is Conceived.
   5.21. Question Types.
   5.3. Comment.
   5.4. How Communicative Skills are Conceived.
   5.5. Marking Criteria.
   5.6. Comment.


   7.1. Face validity.
   7.2. Developing existing sections of the AAP.
   7.3. Extending the AAP.
   7.4. Producing Teaching Materials.
   7.5. Concluding comment.

Bibliography.
Development of the Syndicate's Higher Education Aptitude Tests.

by

Alec Fisher

The Syndicate has been developing aptitude tests for use in higher education since 1985, when it began work on the Law Studies Test. This report summarises the present state of this development work, reviews the progress which has been made and outlines a strategy for future research and development.

1. Background

1.1. The Law Studies Test.

In cooperation with Law Schools Admissions Services of the United States, and following a request for such a test from Cambridge college admission tutors, the Syndicate administered the first Law Studies Test (LST) in June 1987. Since then, the test has been administered annually in June (the 1990 test is attached as Appendix A). The LST is intended as an aptitude test for prospective law students and also for sixth formers wishing to know whether they have the kind of ability demanded by law schools. The test has three sections, called Analytical Reasoning, Logical Reasoning and Reading Comprehension. It is not intended to replace traditional indicators of ability such as A level results, headmasters reports and interviews, but rather to provide additional information on abilities not generally reported elsewhere.

1.2. The Academic Aptitude Profile (AAP) and the Higher Education Aptitude Test (HEAT).

In December 1987, building on their experience with the LST, the Syndicate commissioned Alec Fisher, of the University of East Anglia, to investigate the possibility of producing a general aptitude test for admission to higher education. His report, The Higher Studies Test, was published by the Syndicate in 1989. (Its findings are summarised in Fisher's Proposal to Develop a Higher Studies Test and Research into a Higher Studies Test, which are attached as Appendices B and C.) Fisher's report was subsequently used as the basis for a research programme within the Syndicate, which has lead to the development of the present Academic Aptitude Profile (AAP) and the Higher Education Aptitude Test (HEAT).

The Higher Studies Test was conceived partly as a guidance instrument and partly as a selection instrument, i.e., it aimed
to guide candidates who were seeking admission to higher education concerning their readiness for various areas of academic work. But it also aimed to help admissions tutors of universities, polytechnics and similar degree-level institutions, to select their students. Furthermore, it was intended to be useful both for 'standard' entrants to higher education (school pupils aged 17+ who are taking A-levels) and for 'non-standard' entrants (mature students, access students, and overseas students).

Initial consultations with institutions of higher education suggested a need to separate these functions and to treat these different groups of candidates differently. Furthermore, since a substantial proportion of the pilot institutions indicated a strong interest in a guidance instrument for non-standard entrants, this was given priority during the period 1990 - 1992, and this has led to the development of the present Academic Aptitude Profile. The AAP is seen as having a diagnostic and guidance function, mainly in connection with the admission and progress of non-standard entrants to higher education. It contains three sections, which assess 'Logical Reasoning' ability, 'Mathematical Reasoning' ability, and 'Communicative Skills', respectively.

The selection function which was envisaged for the Higher Studies Test has been taken over by the Higher Education Aptitude Test. Its development was delayed whilst work progressed on the AAP, but it is now back on the agenda. The HEAT is seen as a selection instrument, functioning as part of the admission process to higher education. It also has three sections, and these assess 'Logical Reasoning' ability, 'Mathematical Reasoning' ability, and 'Reading Comprehension' respectively.

1.3. The Assumptions Behind Aptitude Testing.

Before we look in detail at the different sections of the AAP and the HEAT, it is important to be clear about the rationale for such aptitude tests. There are many aptitude tests in use in North America as part of the admission process and elsewhere in higher education. Fisher reported on 5 of these in The Higher Studies Test, the two most important for our purposes being the Scholastic Aptitude Test (SAT) and the Law Schools Admission Test (LSAT).

The rationale behind the SAT is worth spelling out. The main assumption behind its use is that we can identify certain basic 'developed abilities', namely 'verbal reasoning ability' and 'mathematical reasoning ability', which are part of every student's intellectual equipment, which develop slowly over the years, and which are relatively independent of the curriculum being currently followed. It is further assumed that these are the general academic skills necessary for success in college work (regardless of the particular subjects studied). Thus, the SAT
is designed to help to identify those individuals who possess the general (transferable) academic skills necessary for successful college work. It is used in conjunction with High School grades to determine admission to higher education (i.e., it supplies 'supplementary' evidence to 'achievement test' results). Given its purpose, the evidence in favour of using the SAT has to be mainly in terms of its predictive validity (though construct validity is obviously necessary too).

Similar assumptions lie behind the use of the LSAT, and the other North American aptitude tests. Of course, the constructs which are important differ to some extent for different areas of study (law has different requirements from medicine), but the rationale is the same and validation is in terms of predictive validity.

1.4. How the AAP and HEAT differ from North American tests.

In general respects, the assumptions underlying the AAP and the HEAT are similar to those of North American aptitude tests. However, the AAP and the HEAT do not employ the SAT's notion of 'verbal reasoning'; instead, the constructs which are assessed are 'logical reasoning' and 'communicative skills', in the case of the AAP, or 'logical reasoning' and 'reading comprehension', in the case of the HEAT. This difference is due to the fact that the AAP and HEAT are strongly influenced by modern ideas about critical thinking in education (see Fisher 1989, especially chs. 5 and 6).

This influence is partly responsible for another difference between most North American aptitude tests and our instruments. We are especially interested in assessing candidates' critical thinking skills because we think they are of particular importance for subsequent success in higher education. But such skills are commonly thought of as being directly teachable. We do not think of them as 'intrinsic' abilities, like I.Q., but as 'developed abilities' which are directly teachable. However, since we think they are only teachable over the long term, and North American tests assess skills which are also thought of as developing over the long term (though indirectly and as a result of other studies) this may not be a major difference.

The final difference between our instruments and North American aptitude tests concerns their validation. Because the British educational context is so different from the North American one, it is virtually impossible to envisage validating the AAP or the HEAT mainly in terms of predictive validity on the North American model. The problems are fully explained in Fisher Proposal to Develop a Higher Studies Test section III. Our solution is to validate our instruments in terms of 'intrinsic validity', which is a theoretically rich combination of 'face-validity', 'construct validity', and predictive validity; the rationale for doing this is explained in the paper 'Intrinsic Validity and
Usefulness of Profile and Test’ by Fisher, Pollitt and Hamilton. which is attached as Appendix D.

Having described the background to the development of the AAP and the HEAT, we now proceed to look at their contents in detail. We shall first explain how logical reasoning, mathematical reasoning, and communicative skills/reading comprehension are conceived at present. We shall then explore the relationships which exist between the various elements in the AAP and the HEAT, and finally we shall attempt to point the way forward. We begin with logical reasoning.

2. Logical Reasoning.

2.1. Origins.

The initial model for this section was the ‘logical reasoning’ section of the North American Law Schools Admission Test (LSAT). Although the LSAT has been in use for nearly fifty years, it has not always included a ‘logical reasoning’ section: this was first introduced in 1975. The important point about its introduction into the LSAT is that the consultant who advised Educational Testing Service (ETS) on its design was Michael Scriven, arguably the founder of the modern informal logic and critical thinking movement (see Scriven 1976). Ideas very like Scriven’s still inform the design of the logical reasoning section in both the HEAT and the AAP. Indeed the current specifications for the logical reasoning sections of the AAP and HEAT were derived from those for the LSAT, though they have been developed and improved.

2.2. How Logical Reasoning is Conceived.

The intention is that logical reasoning items should test the kinds of reasoning skills which are used in the analysis and criticism of real everyday arguments (i.e., arguments which are or have been actually used by authors with a view to persuading their readers). Logical reasoning items then are expressed in natural language and do not use symbolic languages (or symbolic logic). Stimulus passages either contain some reasoning or they contain sufficient information to serve as a basis for argument. The subject matter of logical reasoning items ranges very widely, and may include anything from cigarette smoking to car maintenance, or from natural science to law. Typical sources for stimulus material include letters to the editor, newspaper and magazine articles and academic articles and books.

2.21. Question types.

The original Higher Studies Test proposal (Fisher 1989) was that the logical reasoning section should test such reasoning skills as the ability.
(i) to identify and clarify critical terms and definitions.
(ii) to identify the reasons and conclusions of a piece of reasoning.
(iii) to identify assumptions underlying reasoning.
(iv) to judge whether an argument is sound.
(v) to see what would strengthen or weaken an argument.
(vi) to draw conclusions from given information.
(vii) to detect reasoning errors.
(viii) to identify parallel arguments.
(ix) to judge whether observations and authorities are reliable.

Logical reasoning items are currently produced by reference to the Law Studies Test: Guide to Writing and Editing Logical Reasoning Items Draft copy 12/02/90. UCLES (hereafter referred to as the LR Guide) and this includes the following item categories.

(i) summarising the main conclusion of an argument
(ii) drawing a conclusion when premises are given
(iii) identifying assumptions
(iv) assessing the impact of additional information
(v) detecting reasoning errors
(vi) identifying techniques used in making an argument
(vii) (a) identifying terms which are critical to the argument and require clarification
     (b) clarifying the meaning of critical terms/phrases
(viii) matching one argument with a second which has the same logical structure
(ix) applying principles

Items of category (iv) - assessing the impact of additional evidence - deal with questions about what would strengthen or weaken an argument. Items of category (vi) concern techniques of argument like arguing by analogy, arguing by reductio ad absurdum, applying a generalisation to particular cases, etc. Items of categories (viii) and (ix) concern parallel/similar arguments.

Examples of items in all these categories are given in the LR Guide. Further examples are contained in the logical reasoning section of the Pilot AAP which is attached as Appendix E (their category is usually easily identifiable by the question they ask - their 'stem'.)

2.3. Comment.

The logical reasoning sections of the AAP and the HEAT are identically conceived, though the average level of item difficulty is likely to be higher in the HEAT.

2.31. Face Validity. One of the most important things to say about the logical reasoning items is that in trialling this material with various institutions of higher education we have
almost universally received the response that it assesses a skill that teachers in such institutions want their students to have (and which they believe is important to subsequent success in higher education); the material has an obvious face validity for them. Furthermore, statistical and other evidence strongly suggests that we have identified a well-defined psychological construct. These views are confirmed by similar development work which is taking place in connection with several North American aptitude tests (these developments are similarly informed by current work on informal reasoning and critical thinking).

2.32. Judging Reliability of Evidence. However, the test of logical reasoning as currently realised is limited in certain respects. For example, it contains no test of the ability to judge whether observations and authorities are reliable; this is now regarded as an important and integral part of the concept of informal reasoning which lies behind tests of logical reasoning. Whether we should attempt to include items which test this skill in the AAP or the HEAT may depend on the use to which the instruments are to be put. If their use is mainly predictive or to give overall 'guidance' concerning a student's placing on a course or remedial needs, what we have may be a sufficient sample of the skills involved. If what is needed is a comprehensive test of logical reasoning skill, or detailed diagnosis of a student's strengths and weaknesses in this domain, then there is a good case for including such items. It is my view that such items should be included in the AAP.

2.33. 'Productive' Logical Reasoning Capacities. The test of logical reasoning is also limited by its multiple-choice format. It does not test candidates' abilities to produce reasoning of their own - what cognitive psychologists call 'productive' capacities. Whether this matters, again depends on the use to which such a test is put. Since the original brief for this development sought an objectively marked test which would be validated mainly in terms of predictive validity, such an instrument might work very well without a test of productive capacities (indeed, this is the case with the LSAT and many similar tests in North America).

In the course of piloting our present test materials, many teachers in institutions of higher education have welcomed the test of logical reasoning as it stands, but have added that a test of these 'productive capacities' would also be very welcome. We suspect that this could be designed using stimulus material similar to the present stimulus material. Of course it could not be objectively marked, and it would need considerable development work to produce mark-schemes for such items which could be used with any reliability. However, this is an interesting possibility which should be explored. As we have argued before (Fisher 1989 p.130), open-ended questioning would probably help with the writing of multiple-choice items, but, much more importantly, it is likely in the British context that open-ended questions would be very attractive as a test of academic
potential. Furthermore, this would move our assessment from a focus on relatively passive skills of logical analysis to one including more active (productive) critical thinking skills.

2.34. Dispositions. Finally, it is widely recognised that someone might be good at answering logical reasoning questions of the kind we are using, without having much inclination to use that skill in the course of thinking about other things. Many teachers in higher education have remarked that they would like to identify students who not only have this skill, but who are also disposed to be critical and questioning. Our present instruments make no attempt to measure whether someone has a questioning and critical disposition, and this is a further avenue worth exploring in view of the importance attached to it by teachers.


3.1. Origin and Purpose.

Most degree courses which have a mathematical content require GCSE or A level mathematics. Such courses often require the content of GCSE or A level mathematics, but in many cases the GCSE or A level pass is required more as an indicator of mathematical ability/aptitude. This is probably so in some of the social, biological and environmental sciences. The mathematical section of the AAP and the HEAT are intended to serve the needs of just such courses (and are not seen as rivals to GCSE or A level mathematics).

It appears that little previous work has been done in Britain on assessing mathematical aptitude at this level, and the only serious attempt, the Test of Academic Aptitude was largely unsuccessful. In arriving at his Higher Studies Test (1989) proposals Fisher reviewed six possible models (four North American, one British, and one Australian); these were the mathematical sections of the Scholastic Aptitude Test, the Test of Academic Aptitude (British), the Graduate Management Admission Test, the Graduate Record Examination, the Commonwealth (Senior) Secondary Scholarships Examination (Australian), and the analytical reasoning section of the Law School Admission Test. He also discussed his proposals with the Shell Centre for Mathematics in Nottingham University. The relevant chapter of Fisher (1989) is attached as Appendix F.

3.2. How Mathematical Reasoning is Conceived.

Fisher (1989) contained only outline proposals for a test of mathematical reasoning ability, because of the many outstanding problems. The intention was that items should test the ability of candidates to understand, interpret and use mathematical information to solve mathematical problems. It was also felt
that though the stimulus material for mathematical reasoning questions could use carefully selected standard mathematical notation, most would not do so, and would present information in a form which required interpretation, using ordinary language, graphs, tables, diagrams or invented notations. However, it was required that all questions based on a given stimulus must be answerable from the information in that stimulus using deductive reasoning alone.

3.21 Question types.

The Fisher (1989) proposal was that candidates should demonstrate such comprehension, reasoning and problem-solving skills as the ability.

(i) to perform elementary arithmetic, algebraic, geometrical and statistical computations,
(ii) to understand and interpret material with a mathematical content,
(iii) to evaluate reasoning about mathematical information and concepts,
(iv) to handle new concepts (mathematical definitions),
(v) to interpret symbolic languages (including new ones),
(vi) to devise suitable symbolisms for employing mathematical techniques,
(vii) to solve problems using mathematical techniques,
(viii) to reason deductively about spatial, temporal, ordering and other mathematical relationships.

Fisher suggested that it should include quantitative, comparison, analytical reasoning, data interpretation and problem-solving questions (for examples of these see Appendix F pp. ), and that it should also apply ideas on critical thinking to the mathematical domain. It was stressed that these were only tentative proposals at this stage, and were intended to serve as a basis for initial consultations with institutions of higher education.


The development of the mathematical reasoning section of the AAP has been undertaken by a team lead by John Hamilton, and consisting of ....... This team has produced the mathematical reasoning items which are to be found in the Pilot AAP which is attached as Appendix E. Mathematical reasoning items are currently produced by reference to an Academic Aptitude Profile: Guide to Writing and Editing Mathematical Reasoning Items (hereafter the MR Guide). This Guide has also been produced by the team mentioned above; it has undergone several revisions and is being revised at present by Alison Green. All subsequent references are to the version due to Alison Green dated 19/3/92.

The present MR Guide contains a brief account of Sternberg's (1985) triarchic theory of intelligence as background to its
conception of mathematical reasoning ability. It claims that we are mainly interested in "productive" or "insightful" thinking (terms derived from Wertheimer 1945). i.e., mathematical problem solving which does not merely require the routine application of familiar knowledge, algorithms or schemas, but which requires a measure of interpretation in handling the data, insight in finding a method, and imagination in finding a solution. It argues that "We are concerned with an individual's capacity to learn. For this reason, it is particularly important that the items are novel and that solutions for them [require insight etc.""] (p.8).

In the light of these ideas, the present MR Guide categorises problems according to whether the "insight" they require involves selective encoding, selective combination, and/or selective comparison:

"Selective encoding means the relevant information is sifted from irrelevant information. For instance, a doctor presented with a patient's case study must determine which facts are relevant for diagnosis or treatment. Selective combination means that certain pieces of information are combined in a relevant way. For instance, having identified which facts are relevant, a doctor must combine this information in a particular way in order to make a diagnosis. Finally, selective comparison means that new information is compared with information previously acquired. Analogical problem solving is one instance of this." (p.9)

Items should require one or more of these processes for solution, and are differentiated, according to the sorts of mathematics involved, into the following categories (p.15).

1. Spatial
   2-dimensions
   3-dimensions
   transformations

2. Data handling
   probility
   data tables
   graphs

3. Numerical
   computation
   enumeration/searching

Examples of items in all these categories are given in the MR Guide. Further examples are contained in the mathematical reasoning section of the Pilot AAP which is attached as Appendix E.
3.3. **Comment.**

3.31. **Testing Aptitude or Knowledge?** All reputable tests of mathematical aptitude recognise that their problem is to avoid the inevitable tendency to test mathematical knowledge/achievement. They usually attempt to do this by reducing the mathematical knowledge assumed by their questions as far as possible, by giving intending candidates advance notice of the mathematical requirements of the test, and by supplying certain basic definitions, formulas and principles which are needed in the test, with the test paper itself (see, for example, NFER 1976 p.126. and current literature on the SAT, the GMAT and similar tests).

Whether this is an adequate procedure depends on the use to which the test is put. The SAT is used in conjunction with High School grades to determine university admission and, increasingly, to guide students in their choice of university courses. Predictive validity results seem to vindicate its use in the first of these roles (though relatively little is known about its use as a guidance instrument). The GMAT, GRE, and similar tests are also used in conjunction with undergraduate grade point averages to determine admission to graduate school courses and, again, predictive validity results suggest that they are successful in this role.

3.32. **The Assumptions Behind Aptitude Testing.** The rationale for using a test of mathematical aptitude rests on the following assumptions: (i) that the needs of the courses for which candidates are applying have been correctly identified, (both in terms of the mathematical knowledge and in terms of the mathematical 'insight/aptitude' required), (ii) that items can be designed to test these, and (iii) that candidates are practiced enough at the activities involved to have reached a stable level of performance (cf. Fisher 1989 ch.3, esp. sect. 3.). Essentially, this requires that what is measured by the test is arrived at by working 'backwards' from what students have to do subsequently in their courses of study and that items are designed which require much the same skills. The validity of such a test will then be partly a matter of its 'face validity' among teachers on the courses concerned, partly a matter of construct validity and partly a matter of predictive validity, depending on the use to which it is put. However, the important point in our present context is that it will not matter if students do poorly on the test due to lack of knowledge, since, by assumption (i) above, that will be the knowledge required for success on the course - and either good predictive validity results will vindicate its use in this role or students will be guided into the remedial courses they need.

Though North American work on aptitude testing has been an important influence on our thinking, the British educational context is different in many ways, and a great deal of the development work which has taken place on the AAP during the past
two years has aimed at designing an instrument which responds to the needs articulated by teachers and admissions officers in higher education institutions in Britain. As we explained earlier, since many of the institutions initially consulted indicated a strong interest in a guidance instrument, this development was given priority, and the AAP is now seen as an aptitude test which is mainly used in a diagnostic and guidance role with non-standard entrants to higher education.

3.33. 'Productive' versus 'Schema-driven' Thinking. The development of the mathematical reasoning section of the AAP has been accompanied by a series of papers, mainly from John Hamilton, which attempt to articulate a difference between mathematical reasoning which involves the relatively routine application of mathematical knowledge, and mathematical reasoning which shows 'mathematical insight', 'mathematical imagination' and hence an 'aptitude' for mathematical work and a capacity to study mathematical material. This has been done in response both to the recognised problem of testing mathematical knowledge/achievement instead of 'aptitude' and in response to the demand for a 'guidance' and 'diagnostic' instrument. Hamilton's latest statement of this distinction is as follows.

"... the Academic Aptitude Profile aims to assess "productive" or "search-driven" thinking rather than "reproductive" or "schema-driven" thinking. In this way, those with the ability to visualize the rearrangement of data and with the ability to deal with situations imaginatively will have an advantage, whatever their mathematical background, over those who are tied to the use of rules in familiar situations. The underlying belief is that productive thinking is more indicative of transference and general abilities while rule-dominated or reproductive thinking is more indicative of specific skills or competence." (JSH/JC 15 January 1992; attached as Appendix G)

It is not easy to test for 'productive' or 'insightful' mathematical thinking - what requires imagination for one person merely requires the application of a familiar method for another. At the same time, Hamilton is attempting to articulate something which is undoubtedly widely felt among teachers in higher education, namely that there is an important difference between being able to apply previously learned routines, and having an aptitude for handling 'mathematical' problems or being 'insightful', and that this difference exists whatever one's level of mathematical knowledge: it seems intuitively clear, for example, that a candidate who wishes to study engineering, might do well at A level mathematics (having been well-prepared) but might score poorly on a test of mathematical aptitude which takes his or her knowledge level for granted but which requires the candidate to be mathematically imaginative in ways appropriate for
engineering. The problem is to make sense of this idea and to find a way of assessing 'productive' mathematical thinking.

On Hamilton's account, to assess mathematical aptitude one must distinguish between measuring mathematical knowledge and measuring mathematical 'insight'. and this suggests that the solution might be a two-part test, the first part testing knowledge and the second part testing 'insight' or 'productive' thinking (indeed Hamilton suggests something like this (ibid.).)

In fact, having worked carefully through all the mathematical reasoning items in the current pilot material (see Appendix) it seems clear to me that the mathematical item writing team has produced a well-designed traditional aptitude test, which could meet the conditions for the proper use of an aptitude test (this is partly a matter for consultation with institutions of higher education as to its face validity, and partly a matter of its predictive validity), which does well in not testing for mathematical knowledge beyond a bare minimum, and which is well-conceived for its intended population, i.e., "non-standard entrants to higher education". However, I do not see that this is a test of "productive" or "insightful" thinking, nor do I see that the attempt to spell out this idea with reference to Sternberg's triarchic theory of intelligence, in terms of 'selective encoding', 'selective combination' and 'selective comparison' is helpful. To see why this is so, it is necessary to work through the items in the present Pilot test in some detail and to attempt to apply the notions of 'productive thinking', 'selective encoding', etc. The present writer's detailed responses are included in Appendix H.

3.34. General Conclusions. In short, these responses lead to the following general conclusions. With a few exceptions, it seems likely that these are good items to test the mathematical aptitude of 'non-standard entrants to higher education'. However, although it is sometimes possible to find clear examples of 'selective encoding', of 'selective combination' and of 'productive thinking', for the most part these are not helpful categories. What really matters is that the likely users of the AAP recognise the items as testing the right skills, that teachers see them as testing the skills they want their students to have, and that candidates recognise their relevance too. This seems to be very likely for most of the items.

On the other hand, the categorisation of items into 'spatial, data handling, numerical', etc., does seem to be very natural and applicable. It might also facilitate the use of part of this section of the test as a diagnostic instrument after suitable refinement. (Alternatively, it might serve as the basis for tests of 'spatial reasoning' ability, as requested by architects.)
The main area of mathematical reasoning skill which is not tested in the present pilot material, but which are included in Fisher's original list above are.

(iv) to handle new concepts (mathematical definitions).
(v) to interpret symbolic languages (including new ones).
(vi) to devise suitable symbolisms for employing mathematical techniques.

The Shell Centre for Mathematics recommended including these skills as important components in mathematical reasoning skill, and it is suggested that the item writing team should consider writing items to test these skills (they may be particularly important for candidates who are planning to work with artificial languages, in computing, etc.)

One concluding remark is necessary. North American aptitude tests are used to supply supplementary information, which is combined with achievement test grades to predict subsequent performance. The AAP is intended mainly for non-standard entrants to higher education. With such candidates, the AAP will not usually be supplying supplementary information and this will compound the problems associated with establishing the 'intrinsic validity' of the instrument. However, if the AAP has good 'intrinsic validity' with standard entrants to higher education, this fact might support its use in a mainly diagnostic and guidance role with non-standard entrants.


The logical reasoning section of the AAP tests skill in everyday, non-deductive reasoning. The mathematical reasoning section tests deductive mathematical skills.

We have spoken of 'productive' skills in both contexts. It is important to note that the term has been used in two very different ways in the two contexts. In the case of logical reasoning, when we spoke of 'productive capacities' we referred to candidates' abilities to produce arguments of their own, rather than to choose from a list, as they do in the case of multiple-choice questions (see p.6 above).

By 'productive thinking' in the mathematical reasoning context was meant "problem solving which does not merely require the routine application of familiar knowledge, algorithms or schemas, but which requires a measure of interpretation in handling the data, insight in finding a method, and imagination in finding a solution" (p.8 above).

None of the mathematical reasoning items is 'productive' in the first sense (because they are all multiple choice), however mathematical reasoning items do require the candidate to think
the problem through first and then to choose an answer. Whereas in the case of logical reasoning items, the distractors are an essential part of what has to be considered in arriving at one's answer. This marks an interesting difference between the ways in which the two kinds of multiple choice items work. Its implications should be considered further; it may mean that a test of 'productive' (in the first sense) logical reasoning capacities is more pressing than a similar test of mathematical reasoning skill, given our overall purposes.

Of course, though none of the logical reasoning items is 'productive' in the first sense (because they are all multiple-choice), they all test 'productive' thinking in the second sense - because they are not "schema-driven" as, say, a test of syllogistic reasoning skill could be.

It is interesting to review the logical reasoning items with Sternberg's 'triarchic theory of intelligence' in mind. For the details of the present writers attempt to do this see Appendix H. In short, although it is clear that these items test 'productive thinking' in the sense that they are not schema-driven, I do not see that Sternberg's categories apply in any helpful way to the analysis of logical reasoning and I think it is better to continue to conceptualise it solely in terms of 'finding assumptions', finding parallel arguments' etc.

In general, the logical reasoning and mathematical reasoning sections of the AAP appear to be testing coherent and well-defined constructs, which measure different skills, and which have an obviously important place in any such instrument.

5. Reading Comprehension and Communicative Skills.

5.1. Origins. Fisher (1989) proposed that the Higher Studies Test should contain a test of reading comprehension very similar to that which was included in the Law Studies Test (LST) which derived in turn from the North American Law Schools Admission Test (LSAT).

However, following initial consultations with institutions of higher education, which showed that there was a particular interest in a guidance instrument for non-standard entrants to higher education - and one which enabled candidates to show what they could do rather than presenting them with a hurdle to jump - it was decided to replace the reading comprehension proposals of the Higher Studies Test with a test of 'communicative skills' in the Academic Aptitude Profile, and the development of this test had had priority during the period 1990 - 1992.

In what follows we first explain the current state of development with respect to 'reading comprehension', and then do the same for 'communicative skills'.

14
5.2. How Reading Comprehension is Conceived.

The LSAT presents itself as a test of reasoning skills and, for this reason, the questions in the reading comprehension section of the LSAT tend to focus on understanding and interpreting the argument or analysis contained in lengthy stimulus passages (four or five passages of about 450 words each). Typically the candidate is asked to identify the main purpose of the passage, what is explicitly stated, what is implied or suggested by the author, what is the author's reasoning or persuasive technique, possible applications of the author's ideas to other situations, and the like (see the reading comprehension section of the 1991 LST which is contained in Appendix A).

5.2.1. Question Types.

Because the Higher Studies Test was also thought of as a test of reasoning skills, especially of critical thinking or higher-order reasoning skills, its reading comprehension section was similar to that of the LSAT and the question types proposed for it were as follows.

(i) clarifying the author's meaning,
(ii) summarising a passage,
(iii) identifying the author's main idea or the primary purpose of a passage,
(iv) identifying information which is explicitly stated in a passage,
(v) applying ideas in a passage to other, similar situations,
(vi) drawing inferences,
(vii) deciding how additional information would influence the author's thesis,
(viii) extending content,
(ix) identifying implicit ideas, etc.

Although no development work has taken place directly on the HEAT during the past two years (because the AAP has had priority), work has been done on the reading comprehension section of the Law Studies Test (LST), and reading comprehension items for the LST are currently produced by reference to the Syndicate's Law Studies Test: Guide to Writing and Editing Reading Comprehension Items (hereafter referred to as the RC Guide). This guide identifies the following question types for the LST reading comprehension section.

(i) Summarising
(ii) Identifying the author's purpose
(iii) Clarifying meaning
(iv) Recognising material in the passage
(v) Organising information
(vi) Inferring author's views
(vii) Inferring views of another
(viii) Identifying purpose of reference
(ix) Describing the author's attitude
(x) Identifying the source of the passage
(xi) Identifying the context of the passage
(xii) Deciding how additional evidence would influence the author's thesis
(xiii) Choosing an analogous situation
(xiv) Extending content

Furthermore, the intention is that

"approximately half the items in the Reading Comprehension section of the test should come from categories 1 to 5, i.e., items that test the candidate's ability to deal with specific information found within the passage and the organisation of that information. The remainder should be drawn from categories 6 to 14, i.e., items that require the candidate to go beyond the information specifically contained in the passage. The candidate must make inferences and draw conclusions based on evidence contained in the passage, but the correct responses cannot be found explicitly in the text" (RC Guide p.21.)

The item types (i) - (xiv) are explained more fully in the RC Guide and examples are provided in all categories.

5.3 Comment

There are many possible conceptions of reading comprehension. The original proposal for the Higher Studies Test focussed on the reasoning skills involved in reading comprehension (as the LSAT does). This focus was deliberate, and was due to the belief that critical thinking or higher level reasoning skills are particularly important in degree-level work. It was also felt that this section of the test could be complemented (especially for overseas students) by a general measure of reading comprehension which could be supplied by the Syndicate's English Language Testing Service.

Although the language of the RC Guide is in many ways similar to that of the Higher Studies Test proposals, the underlying conceptions seem to have moved away from a 'logical/critical thinking' conception to a more 'literary' conception of comprehension and inference. The illustrative passages used in the RC Guide are less obviously 'argumentative' than their counterparts in the Law Schools Admission Test, and their associated questions are less obviously focussed on 'inference' in anything like a logical sense but more on 'inference' in the sense of 'suggests', 'makes one think of', etc.

There appears to have been a subtle shift in the conception of reading comprehension. This may be a good thing, because the obvious question about the Law Schools Admission Test, and the Higher Studies Test proposals which were so closely based upon them, is why one needs both a test of 'logical reasoning' (as
described above p.4ff.) and a test of reading comprehension which is also mainly conceived as a test of reasoning skills. Even though the passages used in reading comprehension exercises are different in style and much longer than logical reasoning passages, there is an obvious overlap in how these two constructs are conceived in the Higher Studies Test proposals. This overlap should be reviewed, and the shift which appears to have taken place in our thinking about reading comprehension should probably be endorsed as making both the reading comprehension and the logical reasoning sections more distinctive and more useful. This review would probably be best conducted by members from both the logical reasoning and the reading comprehension item writing teams.

5.4. How Communicative Skills are Conceived.

Specifications and sample items were first used in a round of consultations with institutions of higher education in the spring of 1990. As a result of these and subsequent discussions, several revisions have taken place, and the following comments are based largely on the Higher Education Guidance Profile: Communicative Skills Component. Draft Specifications (January 1991), which is referred to subsequently as CS Doc and is attached as Appendix J. This document contains the most recent specifications and sample item for the communicative skills component of the Academic Aptitude Profile.

The CS Doc contains the following account of communicative skills.

"The communicative skills component is intended to assess a student’s current ability to read and write, in such a way as to indicate a readiness for academic reading and writing and to indicate areas in which support (in the form of a Communicative Skills course, for example) would be of benefit.

The emphasis is on the ability to write, and it is the importance attached to productive rather than receptive skills which is the most obvious difference between the communicative skills component and the logical reasoning component.

... As far as the Communicative Skills component is concerned, it could be instructive to have, in addition to a score at the component level, a profile of scores for the writing task which, it is planned, would be assessed according to several criteria such as accuracy, organisation, communicative effectiveness etc. A profile at this level would allow differing decisions to be made according to the different types of writing required on different courses of study."

The intention is that stimulus material will be presented to the candidate and he or she will have to produce a single piece of
writing in response to that material. The stimulus will be in the form of one or more pieces of writing (or possibly audio or video material) which deal with a single theme from different perspectives (this is called the 'input'). The piece of writing which is produced (called the 'output') will require the candidate to 'take account of the stimulus material, by explaining it or evaluating it, and to extend it by introducing further ideas or examples and drawing a conclusion'.

Although there will be some comprehension questions to help the candidate with the writing task, and to help the marker assess the candidate's writing, these will not be marked separately. It is the piece of writing, as a 'product', which is being evaluated.

The sorts of comprehension tasks which candidates may be asked to perform are as follows.

(i) Identifying structure, content, sequence of events and procedures
(ii) Finding main ideas which the writer has attempted to make salient
(iii) Identifying the underlying theme or concept
(iv) Identifying ideas in the text, and relationships between them e.g. probability, solution, cause, effect
(v) Identifying, distinguishing and comparing facts, evidence, opinions, implications, definitions and hypotheses
(vi) Evaluating and challenging evidence
(vii) Formulating an hypothesis from underlying theme, concept and evidence
(viii) Reaching a conclusion by relating supporting evidence to the main idea
(ix) Drawing logical inferences

The writing task is designed to sample the candidate's ability to perform the following tasks.

(i) Organising and presenting data
(ii) Listing the stages of a procedure
(iii) Describing an object or event or sequence of events
(iv) Explaining how something works
(v) Presenting the solution to a problem
(vi) Presenting and justifying an opinion, assessment or hypothesis either directly or by implication
(vii) Comparing and contrasting evidence, opinions, implications and hypotheses
(viii) Arguing a case
(ix) Evaluating and challenging ideas, evidence and argument

What is striking about these lists is the frequency with which they refer to argument, reasoning, and justification of one kind
or another, which suggests that this section too is thought of as being at least substantially concerned with assessing reasoning skills.

5.5. **Marking Criteria.** The criteria for marking the communicative skills component are currently explained as follows.
Further details, and an example, are to be found in the CS Doc. Appendix J.

5.6. Comment.

The communicative skills section of the AAP is conceived quite differently from the other sections. The main difference is that it is not a multiple-choice test, but is a test of writing skill and must therefore be subjectively marked.

Since the original brief which lead to the Higher Studies Test proposals was for an objectively-marked (i.e., multiple-choice) aptitude test on the North American model, it is clear that the proposed communicative skills section of the AAP has moved a long way from that idea.

The Law Schools Admission Test does, of course, include a test of writing skill (see Fisher 1969, ch. 4 esp. p. 59) and the Higher Studies Test proposed something very similar (see Fisher's Proposal to Develop a Higher Studies Test p. 11). However, the LSAT writing section is not marked and does not contribute to the candidate's score. Furthermore, the HST proposals went into no detail about how the writing section should be marked or used.

The present proposals for assessing communicative skills appear to be a response to requests from teachers in higher education for an instrument which assesses 'productive' and 'active' writing skills, as opposed to 'receptive' and 'passive' reading comprehension skills. As such they are very interesting and could be very useful: the relevance to the task to the work of many students in higher education is quite obvious: the 'face validity' of the exercise is unmistakable. However, since the candidate's writing is subjectively marked, it will be necessary to give careful instructions (and training in applying these) if markers are to achieve a reasonable measure of inter-rater reliability. Furthermore, since it is not clear whether the method of scoring is intended to yield one overall score or several (compare pp. 17 and 19 above), it is not obvious how this section of the AAP is to be validated. It is certainly hard to see how it can be validated in terms of 'intrinsic validity' like the rest of the AAP (and this is true whether it yields one score or many), however, this may be because it is really conceived as having a mainly diagnostic use - for which 'intrinsic validity' is not a direct issue - rather than as something close to a traditional aptitude test. Indeed it looks as though this may be the section of the AAP which has taken most seriously the request from institutions of higher education for a 'diagnostic and guidance instrument.'

As we remarked above (p.), there is an overlap between what is tested by the present logical reasoning section and the reading comprehension section of the Law Schools Admission Test. This overlap may have been reduced as a result of the apparent shift which has taken place in the way reading comprehension is conceived by the developers of that section of the Law Studies Test.

The communicative skills section of the AAP is so differently conceived from the logical reasoning section (it is basically designed to have a diagnostic function and is not multiple-choice) that there may appear to be few points of contact. However, it is clear from the list of comprehension skills which may be tested and from the list of writing skills which are being assessed, that reasoning of one kind and another plays a large part in both the comprehension and writing part of the exercise.

This is not really surprising, since much recent research on comprehension has emphasised the role of inference in the process (e.g., Johnson-Laird's work). However, it is interesting that the logical reasoning team and the communicative skills team have found their ideas and language converging in this way, because they have worked largely independently of each other and have worked largely from different backgrounds and perspectives (the informal logic and critical thinking tradition on one hand, and the English language and English Language Testing Service tradition, on the other. This suggests that it would be useful for the two teams to discuss their underlying ideas further, with a view to two possible outcomes, (i) since the communicative skills section is designed to have a diagnostic function and is not multiple-choice, its design may assist the logical reasoning team in designing a test of logical reasoning skill which is similarly diagnostic but which tests 'productive' capacities. (ii) since the informal logic and critical thinking tradition has developed some distinctive ideas about the skills involved in reading, writing, reasoning, etc. and has also developed material for assessing these (by other means besides multiple-choice tests) some input from this tradition might help the communicative skills team to develop its specifications further, either to ensure a common focus on reasoning or to differentiate the tests of logical reasoning and communicative skills (in terms of content as well as form).


7.1 Face validity. The experience of the past two years' work strongly suggests that 'face validity' is of crucial importance for the aptitude tests being developed by the Syndicate. The tests have got to 'look right' to prospective users: they must look as though they are testing something which is important to
academic success but which is not tested elsewhere. This suggests that further development should proceed by identifying likely markets, for example Access students, finding which course such candidates predominantly aim for, and then 'working backwards' from the needs of those courses (as identified with the help of teachers on the courses) to the design of an extended Academic Antitude Profile. For example, it may be that there is a significant market for a test of 'symbolic reasoning' ability (for entrants to computing and information sciences). In short, the Syndicate needs to investigate more fully what higher education needs, using the material it has already produced for consultation purposes.

7.2. Developing existing sections of the AAP. (A) The logical reasoning section of the AAP has good face validity in many likely markets. However, it should be extended to include a test of the credibility and reliability of evidence and authorities. Perhaps even more importantly, there is a considerable demand for a test of 'productive' reasoning and critical thinking skills, and this should be investigated as a matter of priority. There are North American models, but they are limited in conception, so this development would require a good deal of work.

(B) The face validity of the mathematical reasoning section of the AAP (as conceived at present) should now be investigated in likely markets. Furthermore, this investigation should be combined with a study of the mathematical knowledge required in different subject areas (social sciences, biological sciences, etc.) to see how the problems associated with testing for mathematical aptitude can best be overcome. It may be possible to test mathematical knowledge and aptitude in separate tests, or it may be best to simply publicise the knowledge required by a given aptitude test (as the SAT does). The present mathematical reasoning section could yield a test of 'spatial reasoning' if there is a demand for such a test, and should be extended to include a test of 'symbolic reasoning' (of special relevance to the information sciences).

(C) Consideration should be given to producing different tests of reading comprehension and communicative skills for different markets; for example there might be one with an 'argumentative' focus for lawyers, social scientists, etc., and one with a more literary focus (Educational Testing Service is experimenting with just such differentiated tests of reading comprehension). The communicative skills team and the logical reasoning team should discuss further their conceptions of reasoning and comprehension (i) to ensure that different test sections test different aptitudes and (ii) to extend the area in which 'productive' skills are tested (to include logical reasoning and critical thinking).

7.3. Extending the AAP. At present the AAP is thought of as focussing on 'higher-order thinking skills' because of their importance to success in higher education and because they are
not reported on separately by any other instruments. It may be that consultations with institutions of higher education will suggest other elements which could be usefully included in an instrument like the AAP. For example, memory is clearly important to success in academic work. The LSAT has experimented with a test of 'reading recall' which might be of interest; this is a test of the candidate's ability to remember the main points of what has been read. The candidate is given a total of 15 minutes to study three passages (ranging from about 500 to 1,000 words in length) and is told in a general way what type of material will be necessary to remember from each. Then, without referring to the passages, the candidate is asked to answer a number of questions on the basis of what was stated or implied in the passages. The recall period can of course be lengthened in various ways.

7.4. Producing Teaching Materials. Since the skills we aim to assess in the AAP and the HEAT are thought of as being directly teachable, the Syndicate should consider preparing some model teaching materials. This would attract the favourable interest of the educational world in a way which tests do not, and would test the credibility of the ideas underlying our assessment instruments. Since there is still a strong emphasis on 'subject-specialism' among teachers in Britain, the Syndicate should investigate the scope for incorporating the teaching and assessment of these 'transferable' thinking skills into particular disciplines at GCSE and A level. For example, an A level in history could include a section or special paper which aimed to assess appropriate transferable thinking skills. Something similar might be done in some branches of mathematics at GCSE level.

7.5. Concluding comment. The Syndicate's aptitude testing programme began with the development of the Law Studies Test (LST) and has now broadened very considerably to include the Academic Aptitude Profile (AAP). Although North American ideas about aptitude testing were very influential in the design of the LST, the AAP is designed for the British educational context and is rather different. It has been developed in cooperation with a number of institutions of higher education, and has been largely a response to the need for an aptitude test with a diagnostic and guidance function for non-standard entrants to higher education. It seems likely that there are many other opportunities for aptitude testing in the academic world and in the world of employment. Though much work remains to be done, the Syndicate has a good base on which to build.

Bibliography.