Hughes, A. 1990. *Testing for Language Teachers.* Cambridge: Cambridge University Press.

Ingram, E. 1977. Basic Concepts in Testing. In J.P.B. Allen and A. Davies (eds.), *Testing and Experimental Methods.* Oxford: Oxford University Press.

Lord, F.M. 1980. *Applications of Item Response Theory to Practical Testing Problems.* Hillsdale, NJ: Lawrence Erlbaum.

Oller, J.W. 1979. *Language Tests at School.* London: Longman.

Popham, W.J. 1990. *Modern Educational Measurement: A Practitioner's Perspective.* 2nd edition. Boston, Mass.: Allyn and Bacon.

Weir, C.J. 1990. *Communicative Language Testing.* Englewood Cliffs, NJ: Prentice-Hall Regent.

2 Test specifications

The questions that this chapter seeks to answer in detail are: What are test specifications? Who needs test specifications? What should test specifications look like? How can we draw up test specifications? What do current EFL examinations prepare in the way of specifications?

2.1 What are test specifications?

A test's specifications provide the official statement about what the test tests and how it tests it. The specifications are the blueprint to be followed by test and item writers, and they are also essential in the establishment of the test's construct validity.

Deriving from a test's specifications is the test syllabus. Although some UK examination boards use *specifications* and *syllabus* interchangeably, we see a difference between them. A test specification is a detailed document, and is often for internal purposes only. It is sometimes confidential to the examining body. The syllabus is a public document, often much simplified, which indicates to test users what the test will contain. Whereas the test specification is for the test developers and those who need to evaluate whether a test has met its aim, the syllabus is directed more to teachers and students who wish to prepare for the test, to people who need to make decisions on the basis of test scores, and to publishers who wish to produce materials related to the test.

The development and publication of test specifications and syllabuses is, therefore, a central and crucial part of the test construction and evaluation process. This chapter will describe the sorts of things that test specifications and syllabuses ought to contain, and will consider the documents that are currently available for UK EFL tests.

2.2 Who needs test specifications?

As has already been suggested, test specifications are needed by a range

of different people. First and foremost, they are needed by those who produce the test itself. Test constructors need to have clear statements about who the test is aimed at, what its purpose is, what content is to be covered, what methods are to be used, how many papers or sections there are, how long the test takes, and so on. In addition, the specifications will need to be available to those responsible for editing and moderating the work of individual item writers or teams. Such editors may operate in a committee or they may be individual chief examiners or board officials. (See Chapter 3 for further discussion of the editing process.) In smaller institutions, they may simply be fellow teachers who have a responsibility for vetting a test before it is used. The specifications should be consulted when items and tests are reviewed, and therefore need to be clearly written so that they can be referred to easily during debate. For test developers, the specifications document will need to be as detailed as possible, and may even be of a confidential nature, especially if the test is a 'high-stakes' test.

Test specifications are also needed by those responsible for or interested in establishing the test's validity (that is, whether the test tests what it is supposed to test). These people may not be the test constructors, but outsiders or other independent individuals whose needs may be somewhat different from those of the item writers or editors. It may be less important for validators to have 'practical' information, for example, about the length of the test and its sections, and more important to know the theoretical justification for the content: what theories of language and proficiency underpin the test, and *why* the test is the way it is.

Test users also need descriptions of a test's content, and different sorts of users may need somewhat different descriptions. For example, teachers who will be responsible for the learners placed in their classes by a test need to know what the test scores mean: what the particular learners know, what they can do, what they need to learn. Although the interpretation of test scores is partly a function of how scores are calculated and reported (see Chapter 7), an understanding of what scores mean clearly also relates to what the test is testing, and therefore to some form of the specifications.

Teachers who wish to enter their students for some public examination need to know which test will be most appropriate for their learners in relation to the course of instruction that they have been following. They need information which will help them to decide which test to choose from the many available. Again, some form of the specifications will help here – probably the simplified version known as the syllabus.

Admissions officers who have to make a decision on the basis of test scores will also need some description of a test to help them decide whether the test is valid for the particular decisions to be taken: for university admissions purposes, a test that does not measure academic-related language skills is likely to be less valid than one that does.

Finally, test specifications are a valuable source of information for publishers wishing to produce textbooks related to the test: textbook writers will wish to ensure that the practice tests they produce, for example, are of an appropriate level of difficulty, with appropriate content, topics, tasks and so on.

All these users of test specifications may have differing needs, and writers of specifications need to bear the audience in mind when producing or revising their specifications. What is suitable for one audience may be quite unsuitable for another.

2.3   What should test specifications look like?

Since specifications will vary according to audience, this section is divided according to the different groups of people needing specifications. However, as the principal user is probably the test writer/editor, the first section is the longest and encompasses much that might be relevant for other users.

2.3.1   Specifications for test writers

Test writers need guidance on practical matters that will assist test construction. They need answers to a wide range of questions. The answers to these questions may also be used to categorise an item, text or test bank so that once items have been written and pretested, they can be classified according to one or more of the following dimensions, and stored until required.

1. What is the *purpose* of the test? Tests tend to fall into one of the following broad categories: placement, progress, achievement, proficiency, and diagnostic.

   Placement tests are designed to assess students' level of language ability so that they can be placed in the appropriate course or class. Such tests may be based on aspects of the syllabus taught at the institution concerned, or may be based on unrelated material. In some language centres students are placed according to their rank in the test results so that, for example, the students with the top eight scores might go into the top class. In other

centres the students' ability in different skills such as reading and writing may need to be identified. In such a centre a student could conceivably be placed in the top reading class, but in the bottom writing class, or some other combination. In yet other centres the placement test may have the purpose of deciding whether students need any further tuition at all. For example, many universities give overseas students tests at the start of an academic year to discover whether they need tuition in the language or skill used at the university.

Progress tests are given at various stages throughout a language course to see what the students have learnt.

Achievement tests are similar, but tend to be given at the end of the course. The content of both progress and achievement tests is generally based on the course syllabus or the course textbook.

Proficiency tests, on the other hand, are not based on a particular language programme. They are designed to test the ability of students with different language training backgrounds. Some proficiency tests, such as many of those produced by the UK examination boards, are intended to show whether students have reached a given level of general language ability. Others are designed to show whether students have sufficient ability to be able to use a language in some specific area such as medicine, tourism or academic study. Such tests are often called Specific Purposes (SP) tests, and their content is generally based on a *needs analysis* of the kinds of language that are required for the given purpose. For example, a proficiency test for air traffic controllers would be based on the linguistic skills needed in the control tower.

Diagnostic tests seek to identify those areas in which a student needs further help. These tests can be fairly general, and show, for example, whether a student needs particular help with one of the four main language skills; or they can be more specific, seeking perhaps to identify weaknesses in a student's use of grammar. These more specific diagnostic tests are not easy to design since it is difficult to diagnose precisely strengths and weaknesses in the complexities of language ability. For this reason there are very few purely diagnostic tests. However, achievement and proficiency tests are themselves frequently used, albeit unsystematically, for diagnostic purposes.

2. What sort of *learner* will be taking the test – age, sex, level of proficiency/stage of learning, first language, cultural background, country of origin, level and nature of education, reason for taking the test, likely personal and, if applicable, professional interests, likely levels of background (world) knowledge?

3. How many *sections/papers* should the test have, how long should they be and how will they be differentiated – one three-hour exam, five separate two-hour papers, three 45 minute sections, reading tested separately from grammar, listening and writing integrated into one paper, and so on?

4. What *target language situation* is envisaged for the test, and is this to be simulated in some way in the test content and method?

5. What *text types* should be chosen – written and/or spoken? What should be the sources of these, the supposed audience, the topics, the degree of authenticity? How difficult or long should they be? What functions should be embodied in the texts – persuasion, definition, summarising, etc.? How complex should the language be?

6. What *language skills* should be tested? Are enabling/micro skills specified, and should items be designed to test these individually or in some integrated fashion? Are distinctions made between items testing main idea, specific detail, inference?

7. What *language elements* should be tested? Is there a list of grammatical structures/features to be included? Is the lexis specified in some way – frequency lists etc.? Are notions and functions, speech acts or pragmatic features specified?

8. What sort of *tasks* are required – discrete point, integrative, simulated 'authentic', objectively assessable?

9. How many *items* are required for each section? What is the relative weight for each item – equal weighting, extra weighting for more difficult items?

10. What test *methods* are to be used – multiple choice, gap filling, matching, transformation, short answer question, picture description, role play with cue cards, essay, structured writing?

11. What *rubrics* are to be used as instructions for candidates? Will examples be required to help candidates know what is expected? Should the criteria by which candidates will be assessed be included in the rubric?

12. Which *criteria* will be used for assessment by markers? How important is accuracy, appropriacy, spelling, length of utterance/script, etc.?

Some of the above questions inevitably partially cover the same ground: for example 'text type', 'nature of text' and 'complexity of text' all overlap. However, it is nevertheless helpful to address them from a variety of angles. Complete taxonomies for specifications are beyond the scope of this chapter, and in any case it is impossible, given the nature of language and the variety of different tests that can be envisaged, to be exhaustive. A very useful taxonomy that readers might consider, however, is that developed by Lyle Bachman in *Fundamental Considerations in Language Testing* (1990). This is described more fully in the next section, but in order to give the reader an idea of what specifications for test writers might contain, there follows a fictional example of the specifications for a reading test. (For an example of some more detailed specifications for an academic reading test, see Davidson and Lynch 1993.)

TEST OF FRENCH FOR POSTGRADUATE STUDIES

Specifications for the Reading Test

General Statement of Purpose

The Test of French for Postgraduate Studies is a test battery designed to assess the French language proficiency of students who do not have French as their first language and who hope to undertake postgraduate study at universities and colleges where French is the medium of instruction.

The aim of the battery is to select students who have sufficient French to be able to benefit from an advanced course of academic study, and to identify those linguistic areas in which they might need help.

The focus of the test battery is on French for Academic Purposes.

The Test Battery

The battery consists of four tests:

Reading	60 minutes
Writing	60 minutes
Listening	30 minutes
Speaking	15 minutes

Separate scores are reported for the four tests. There is a different set of specifications for each of the four tests.

Reading Test

Time allowed: One hour

Test focus: The level of reading required for this test should be in the region of levels 5 to 7 of the English Speaking Union (ESU) Yardstick Scale.

Candidates will have to demonstrate their ability to read textbooks, learned articles and other sources of information relevant to academic education. Candidates will be expected to show that they can use the following reading skills:

a)   skimming
b)   scanning
c)   getting the gist
d)   distinguishing the main ideas from supporting detail
e)   distinguishing fact from opinion
f)   distinguishing statement from example
g)   deducing implicit ideas and information
h)   deducing the use of unfamiliar words from context
i)   understanding relations within the sentence
j)   understanding relations across sentences and paragraphs
k)   understanding the communicative function of sentences and paragraphs

Source of texts: Academic books, papers, reviews, newspaper articles relating to academic subjects. The texts should not be highly discipline-specific, and should not disadvantage students who are not familiar with the topics. All passages should be understandable by educated readers in all disciplines. A glossary of technical terms should be provided where necessary.

There should be four reading passages, each of which should be based on a different academic discipline. Two of the texts should be from the life and physical sciences, and two from the social sciences. As far as possible the four texts should exemplify different genres. For example, one text might consist of an introduction to an academic paper, and the other three might consist of a review, a description of some results and a discussion.

The texts should be generally interesting, but not distressing. Recent disasters and tragedies should be avoided.

Passages should be based on authentic texts, but may receive minor modifications such as abridgement and the correction of grammatical errors.

The length of the passages together should total 2,500 to 3,000 words.

Test tasks: Each test question should sample one or more of the reading abilities listed above. Test writers should try to achieve a balance so that one or two skills are not over tested at the expense of the others.

Item types: The Reading Test should contain between 40 and 50 items – approximately 12 items for each reading passage. Each reading passage and its items will form one sub-test. Each item will be worth one mark. Items may be open-ended, but they must be objectively markable. Item writers should provide a comprehensive answer key with their draft test.
Item writers should use a variety of item types. These may include the following:

identifying appropriate headings
matching
labelling or completing diagrams, tables, charts, etc.
copying words from the text
information transfer
short answer questions
gap filling
sorting events or procedures into order

Item writers may use other types of test item, but they should ensure that such items are objectively markable.

Rubrics: There is a standard introduction to the Reading Test which appears on the front of each Reading Test question paper. Item writers, however, should provide their own instructions and an example for each set of questions. The language of the instructions should be no higher than Level 4 of the ESU Yardstick Scale.

2.3.2 Specifications for test validators

Every test has a theory behind it: some abstract belief of what language is, what language proficiency consists of, what language learning involves and what language users do with language. This theory may be more or less explicit. Most test constructors would be surprised to hear that they have such a theory, but this does not mean that it is not there, only that it is implicit rather than articulated in metalanguage. Every test is an operationalisation of some beliefs about language,

whether the constructor refers to an explicit model or merely relies upon 'intuition'.

Every theory contains *constructs* (or psychological concepts), which are its principal components and the relationship between these components. For example, some theories of reading state that there are many different constructs involved in reading (skimming, scanning, etc.) and that the constructs are different from one another. Construct validation involves assessing how well a test measures the constructs. For validation purposes, then, test specifications need to make the theoretical framework which underlies the test explicit, and to spell out relationships among its constructs, as well as the relationship between the theory and the purpose for which the test is designed.

The Bachman model mentioned above is one such theoretical framework, which was developed for the purpose of test analysis. It was used by Bachman et al. 1988, for example, to compare tests produced by the University of Cambridge Local Examinations Syndicate (UCLES) and Educational Testing Service (ETS), but it could equally be used as part of the test construction/validation process. The taxonomy is divided into two major sections: communicative language ability and test method facets. The model below shows how each section consists of a number of components.

Bachman's Frameworks of Communicative Language Ability and Test Method Facets

A. COMMUNICATIVE LANGUAGE ABILITY

1. ORGANISATIONAL COMPETENCE

   Grammatical Competence
   Vocabulary, Morphology, Syntax, Phonology/Graphology
   Textual Competence
   Cohesion, Rhetorical organisation

2. PRAGMATIC COMPETENCE

   Illocutionary Competence
   Ideational functions, Manipulative functions, Heuristic functions, Imaginative functions
   Sociolinguistic Competence
   Sensitivity to differences in dialect or variety, Sensitivity to differences in register, Sensitivity to naturalness, Ability to

interpret cultural references and figures of speech

(Bachman 1990: Chapter 4)

B. TEST METHOD FACETS

1. FACETS OF THE TESTING ENVIRONMENT

Familiarity of the Place and Equipment
Personnel
Time of Testing
Physical Conditions

2. FACETS OF THE TEST RUBRIC

Test Organisation
Salience of parts, Sequence of parts, Relative importance of parts
Time Allocation
Instructions
Language (native, target), Channel (aural, visual), Specification of procedures and tasks, Explicitness of criteria for correctness

3. FACETS OF THE INPUT

Format
Channel of presentation, Mode of presentation (receptive), Form of presentation (language, non-language, both), Vehicle of presentation ('live', 'canned', both), Language of presentation (native, target, both), Identification of problem (specific, general), Degree of speededness
Nature of Language
Length, Propositional content (frequency and specialisation of vocabulary, degree of contextualisation, distribution of new information, type of information, topic, genre), Organisational characteristics (grammar, cohesion, rhetorical organisation), Pragmatic characteristics (illocutionary force, sociolinguistic characteristics)

4. FACETS OF THE EXPECTED RESPONSE

Format
Channel, Mode, Type of response, Form of response, Language of response
Nature of Language
Length, Propositional content (vocabulary, degree of contextualisation, distribution of new information, type of information, topic, genre), Organisational characteristics (grammar, cohesion, rhetorical organisation), Pragmatic characteristics (illocutionary force, sociolinguistic characteristics)
Restrictions on Response
Channel, Format, Organisational characteristics, Propositional and illocutionary characteristics, Time or length of response

5. RELATIONSHIP BETWEEN INPUT AND RESPONSE

Reciprocal
Nonreciprocal
Adaptive

(Bachman 1990: 119)

Other models on which test specifications have been based in recent years include *The Council of Europe Threshold Skills*, and Munby's *Communication Needs Processor* (1978), which informed the design and validation of both the Test of English for Educational Purposes (TEEP) by the Associated Examining Board (AEB) and the UCLES/British Council English Language Testing Service (ELTS) test. Other less explicitly articulated models of communicative competence are behind the design if not the validation of tests like the former Royal Society of Arts (RSA) Examination in the Communicative Use of English as a Foreign Language (CUEFL).

The content of test specifications for test validators will obviously depend upon the theoretical framework being used, and will not therefore be dealt with at length here. Nevertheless, the reader should note that much of the content outlined in the previous section would also be included in validation specifications. In particular, information should be offered on what abilities are being measured, and the interrelationships of these abilities, what test methods are to be used and how these methods influence (or not) the measurement of the abilities, and what criteria are used for assessment. Of less relevance to this sort of specification is perhaps the matter of test length, timing, item exemplification, text length, and possibly even difficulty: in short, matters that guide item writers in producing items but which are not known to have a significant effect on the measurement of ability. It should, however, be emphasised at this point that language test researchers are still uncertain as to which variables do affect construct validity and which do not, and the most useful, if not the most practical, advice is that validation specifications should be more, rather than less, complete.

A discussion of the value of any particular model or theory is well beyond the scope of this book, and is properly the domain of books dealing with language, language learning and language use. Nevertheless, any adequate treatment of test design must include reference to relevant theories. For example, *Fundamental Considerations in Language Testing* (Bachman 1990) is essentially a discussion of a model of language, and John Oller's *Language Tests at School* (1979) contains an extended treatment of his theory of a grammar of pragmatic expectancy which provides the rationale for the types of tests he advocates. Sadly, however, too many textbooks for language testers contain little or no discussion of the constructs which are supposedly being tested by the tests and test/item types that are discussed. Yet it is impossible to design, for example, a reading test without some statement of what reading is and what abilities are to be measured by an adequate test of reading. Such a statement, therefore, should also form part of test specifications.

2.3.3 Specifications for test users

Test specifications which are aimed at test users (which we will call *user specifications* for the sake of this discussion, and which include the notion of syllabus presented in Section 1 above) are intended to give users a clear view of what the test measures, and what the test should be used for. They should warn against specific, likely or known misuses.

A typical example of misuse is the attempt to measure students' language progress by giving them the same proficiency test before and after their course. Proficiency tests are such crude measures that if the interval is three months or less there may well be no improvement in the students' scores, and some students' scores may even drop. To avoid such misuse, the specifications should accurately represent the characteristics, usefulness and limitations of the test, and describe the population for which the test is appropriate.

Such user specifications should provide representative examples of item types, or, better, complete tests, including all instructions. They should provide a description of a typical performance at each significant grade or level of the test and, where possible and relevant, a description of what a candidate achieving a pass or a given grade can be expected to be able to do in the real world. In addition to examples of items or tests, it is particularly helpful to teachers but probably also to learners if examples can be provided of candidates' performances on typical or previous items/tests, and a description of how the criteria used to assess those performances apply to the examples.

For many examinations it may also be helpful to provide users with a description of what course of study or what test preparation would be particularly appropriate prior to taking the examination.

It is clearly important that candidates are given adequate information to enable them to know exactly what the test will look like: how long it will be, how difficult it is, what the test methods will be, and any other information that will familiarise them with the test in advance of taking it. The intention of such specifications for candidates should be to ensure that as far as possible, and as far as is consistent with test security, candidates are given enough information to enable them to perform to the best of their ability.

2.4 How can we draw up test specifications?

The purpose for which the test will be used is the normal starting point for designing test specifications. This should be stated as fully as possible. For example:

> Test A is used at the end of the second year of a three-year Bachelor of Education degree course for intending teachers of English as a Foreign Language. It assesses whether students have sufficient competence in English to proceed to teaching practice in the final year of study. Students who fail the test will have an opportunity to re-sit a parallel version two months later. If they subsequently fail, they will have to repeat the second year English course. Although the test relates to the English taught in the first two years, it is a proficiency test, not a measure of achievement, and is not intended to reflect the syllabus.

or:

> Test B is a placement test, designed to place students applying for language courses at the Alliance Française into classes appropriate to their language level.

or:

> Test C is intended to diagnose strengths and weaknesses of fourth year secondary school pupils in German grammar.

From the above examples, it should be clear that the test's purpose will influence its content. Test A will probably need to include measures of abilities that are relevant to the student teachers' use of English in English classes during their teaching practice. Test B may attempt to sample from the syllabus, or achievement tests, of each course level within the Alliance Française. Test C will need to refer to a model of German grammar, a list of structures that students need to know at this

level, and probably to typical problems students have and errors they produce.

Having determined the purpose and the target population, test designers will then need to identify a framework within which the test might be constructed. This may be a linguistic theory – a view of language in the case of proficiency tests or a definition of the components of aptitude in the case of aptitude tests – or it may be considered necessary first to engage in an analysis of target language situations and use, and the performance which the test is intended to predict. In this case, designers may decide to undertake analyses of the likely jobs/tasks which learners may have to carry out in the future, and they may have to undertake or consult analyses of their linguistic needs.

Needs analyses typically involve gathering information on what language will be needed by test candidates in relation to the test's purpose. This might involve direct observation of people in target language use situations, to determine the range of variables relevant to language use. It may involve questionnaires or interviews with language users, or the consultation of relevant literature or of experts on the type of communication involved. An example of the sorts of variables that might be involved have been listed by Munby in his Communication Needs Processor (Munby 1978), and these include:

Participant	age, sex, nationality, domicile
Purposive domain	type of ESP involved, and purposes to which it is to be put
Setting	e.g. place of work, quiet or noisy environment, familiar or unfamiliar surroundings
Interaction	participant's role, i.e. position at work, people with whom he/she will interact, role and social relationships
Instrumentality	medium, mode and channel of communication, e.g. spoken or written communication, monologue or dialogue, textbook or radio report
Dialect	e.g. British or American English
Target Level	required level of English
Communicative Event	e.g. at a macro level: serving customers in restaurant, attending university lectures, and at a micro level, taking a customer's order, introducing a different point of view
Communicative Key	'the tone, manner and spirit in which an act is done' (Hymes 1972).

The literature on English for Specific Purposes (ESP) (see, for example, Hutchinson and Waters 1987; Robinson 1980; Swales 1985) is useful for test developers who need to conduct some form of needs analysis

before they can begin to draw up their specifications. Note that both TEEP and ELTS were initially developed using some form of Munby-style needs analysis.

Needs analyses usually result in a large taxonomy of variables that influence the language that will be needed in the target situation. From this taxonomy, test developers have to sample tasks, texts, settings and so on, in order to arrive at a manageable test design. However, the ELTS Revision Project which was responsible for developing the International English Language Testing System (IELTS) test, successor to the original ELTS, proceeded somewhat differently. Once the main problems with ELTS had been identified (see Criper and Davies 1988), the revision project undertook an extensive data-gathering exercise in which a variety of test users such as administrators, teachers and university officials were asked how they thought the ELTS test should be revised. At the same time the literature relating to English for Academic Purposes (EAP) proficiency testing was reviewed, and eminent applied linguists were asked for their views on the nature of language proficiency and how it should be tested in IELTS. Teams of item writers were then asked to consider the data that had been collected and to produce draft specifications and test items for the different test components. These drafts were shown to language testers and teachers, and also to university lecturers in a wide range of academic disciplines. The lecturers were asked whether the draft specifications and sample texts and tasks were suitable for students in their disciplines, and whether other text types and tasks should be included. The item writers then revised the test battery and its specifications to take account of all the comments. By proceeding in this way the revision project members were able to build on existing needs analysis research, and to carry out a content validation of the draft test (see Alderson and Clapham 1992a and 1992b, and Clapham and Alderson forthcoming). For a discussion of how to develop ESP test specifications, and the relationship between needs analyses, test specifications and informants, see Alderson 1988b.

The development of an achievement test is in theory an easier task, since the language to be tested has been defined, at least in principle, by the syllabus upon which the test will be based. The problem for designers of achievement tests is to ensure that they adequately sample either the syllabus or the textbook in terms of content and method.

Hughes 1988 has argued that while he agrees with the general distinction made between proficiency and end of course achievement tests, he does not agree that different procedures should be followed for deciding their content. He argues that such achievement tests should be based on course *objectives*, rather than course *content*, and would

therefore be similar or even identical to proficiency tests based on those same objectives.

At the end of this chapter there is a checklist containing the possible points to be covered in a set of specifications. This checklist is presented in a linear fashion, but usually the design of a test and its specifications is cyclical, with early drafts and examples being constantly revised to take account of feedback from trials and advisers.

2.5 Survey of EFL Examinations Boards: Questionnaire and Documentation

In this section we describe the EFL examinations boards' approach to test specifications: how they draw them up and what the specifications contain. We shall report the answers to the questionnaire and we shall also, as far as possible, refer to the documents the boards sent us. (See Chapter 1 for details of how this survey was conducted.) This is not always easy, because the boards use different methods and different terminology. For example, few of them use the expression *specifications*: some refer to *syllabuses*, some to *regulations* and some to *handbooks*, and the meaning of each of these terms differs from board to board. In addition, some of the boards' procedures are confidential or are not well publicised. Nor do most of the boards say for whom their publications are intended, so we are not able to consider the documents' intended audiences.

Our report on the boards' responses to this section of the questionnaire is longer than those in later chapters. This reflects the detail of the responses: not only did the boards give their fullest answers to those questions relating to test specifications, but the documents they sent contained a wide variety of information on such aspects of the exams as their aims and syllabuses.

Since UCLES filled in separate questionnaires for each of their EFL exams, it is difficult to combine their results with those from the other boards, where answers sometimes referred to one exam and sometimes to more than one. In addition, subject officers of four of the UCLES exams answered separate questionnaires for individual papers within those exams. The UCLES answers have therefore been somewhat conflated. In Table 2.1, which gives the breakdown of all the boards' answers to Questions 6 to 10, the UCLES figures represent the majority of the answers. If, for example, out of the five papers in an exam, three subject officers said Yes to a question and two said No, the answer is reported as Yes. (For details of the wording of each sub-question, see below, and for a copy of the whole questionnaire, see Appendix 2.)

QUESTIONS 6 TO 7(d): *Does your board publish a description of the content of the examinations(s); does this include a statement of its purpose, and a description of the sort of student for whom it is intended?*

TABLE 2.1 THE EXAMINATIONS BOARDS' ANSWERS

	11 exam boards			8 UCLES exams	
Question	Yes	No	N/A	Yes	No
6. Publish description	11	0	0	8	0
7. Does this include:					
a) purpose	11	0	0	8	0
b) which students	11	0	0	8	0
c) level of difficulty	11	0	0	8	0
d) typical performance	10	1	0	5	3
e) ability in 'real world'	9	1	1	4	4
f) course of study	2	7	1	1	7
g) content of exam					
structures	6	3	0	2	6
vocabulary	5	4	0	2	6
language functions	6	3	0	2	6
topics	6	3	0	3	5
text length	6	2	1	5	2
question types	9	0	0	8	0
question weighting	8	1	0	3	5
timing of papers	9	0	0	8	0
timing of sections	6	3	0	1	7
h) criteria for evaluation	9	1	0	2	6
i) derivation of scores	4	6	0	2	5
j) past papers	8	0	2	6	0
k) past student performance	2	5	2	7	1
8. Needs analysis	7	1	0	4	3
9. Guidance to item writers	7	1	2	8	0

As can be seen from Table 2.1, everyone said Yes to Questions 6 to 7(c). All the boards published descriptions of their examinations, and each description included a statement of the purpose of the exam, a description of the sort of student for whom it was intended and a description of its level of difficulty. A study of the published documents showed that the level of detail, however, varied from board to board. Here are a few examples:

STATEMENT OF PURPOSE

In its syllabus, the Joint Matriculation Board (JMB) gives one of the fullest descriptions of the purpose of an exam: