Exploring task difficulty in EFL reading assessment: The case of multiple matching tasks

Cseresznyés Mária

PhD értekezés

Supervisor: Professor J. Charles Alderson

Doctoral Programme in Applied Linguistics

University of Pécs

2008

Abstract

This thesis investigates the relationship between characteristics of multiple matching reading tasks and learners' performance on EFL reading comprehension tests. Multiple matching tasks are one of the testing techniques most commonly used in recent tests of reading in a second or foreign language. Such tasks are also included in the Hungarian School-leaving Examination as an innovative way of assessing Hungarian students' foreign language reading abilities. However, a review of relevant literature suggests that no previous research has investigated the effect of task and item features specific to such test techniques. In the research reported in this thesis, three studies were conducted to explore the nature of multiple matching task techniques. This research investigated multiple matching reading tasks developed and pre-tested/piloted by the British Council-supported Hungarian English Examinations Reform Project.

Study One used content analysis to identify item characteristics likely to affect performance on the tasks under investigation. Study Two, relying on think-aloud protocols generated by the subjects involved in the research, explored the skills, knowledge and processes students actually used when responding to the tasks and items. Study Three investigated the relationship between the item characteristics identified through content analysis and those revealed by the verbal protocols, on the one hand, and the item characteristics and the empirical item difficulties, on the other.

As a result of the various analyses, nine of the 15 item characteristic variables identified were shown to have important effects on the difficulty of the reading items examined. The study showed that many of the variables underlying performance on the items were identified through content analysis, which suggests that test developers should carefully examine the content of test items. In line with previous research into item difficulty, the study also revealed that, on the one hand, different reading items are differentially difficult for individual test takers and, on the other hand, test takers may arrive at the same correct or incorrect answers using very different processes.

One of the most important findings of the research is that students' verbal reports showed that they often failed to select the correct answer to an item despite demonstrating the skill or knowledge intended to be assessed by the item in question, and that, on the other hand, there were cases where students responded to the item correctly despite an apparent failure to understand the meaning of relevant sections of the reading text.

The general conclusion drawn from the findings of the three studies is that further research using multiple data sources, including content analysis and introspective data, is required to further explore the effects of item characteristic variables on the difficulty of reading test items in general, and the type of reading items investigated in this research, in particular.

Contents

Chapter 1 Introduction

1.1 The purpose of the dissertation

This dissertation investigates the effects of task and item features on learners'
performance on EFL (English as a Foreign Language) reading comprehension tests,
with a focus on characteristics of matching tasks developed by the Hungarian
Examinations Reform Project for purposes of the new Hungarian School-leaving
Examination in English. The rationale for focusing on this type of reading test tasks is
many-fold. First, it represents one of the test methods most commonly used in recent
tests of reading comprehension and, more importantly, such tasks are included in the
new Hungarian Matura examination as an innovative way of assessing school-leavers'
foreign language reading abilities. On the other hand, a review of relevant literature
suggests that, although previous studies have investigated a range of factors underlying
performance on reading tests, they have typically based their investigation on either the
cloze procedure or the traditional 4-option multiple-choice questions, whilst to date no
(published) research has focused on the impact of task and item features specific to
matching tasks. From a different perspective, the vast majority of previous studies have
examined the issue of item difficulty either through relating content characteristics to
item statistics (Freedle and Kostin 1993; Bachman et al. 1996; Buck et al. 1997) or
exploring verbal report data on reading and test taking strategies (Alderson 1990b;
Nikolov 2001a, 2001b; Cohen and Upton 2006), whereas only a few studies have
examined the relationship among all three types of information, that is, item
characteristics, the skills, processes and strategies used by test takers in actual
completion of the items, and item difficulty (Anderson et al. 1991; Jang 2005), which

approach is used to explore characteristics of the reading items under investigation in this dissertation.

In the design and development of language tests, it is considered to be a major concern to determine the extent to which scores on a test reflect factors other than the language abilities we want to measure and minimize unwanted variation in the scores due to the effects of test method facets (Bachman 1990). However, to be able to control for the influences of factors that may introduce construct irrelevant variance into test scores, in practical terms, to be able to design tests that, discriminating better readers from poorer readers, fairly reflect the reading abilities of the test taker, it is crucial for test constructors to be aware of variables affecting the difficulty of their tasks and items. The principal aims of this research are to

- describe the content of the matching tasks and items under investigation and identify item characteristic variables likely to affect performance on these reading items (Study One),

- explore the skills, knowledge and processes that test takers actually use, and the difficulties they encounter when completing the items (Study Two) and

- examine whether the item characteristics identified in Study One and Study Two relate to the empirical difficulty of the items (Study Three).

Study One employs the methodology of content analysis to determine variables that may impact on the difficulty of the reading items focused on in this research. Study Two investigates the same issue from the perspective of the test taker, and applies think-aloud protocols generated by the subjects participating in the study in order to examine how test takers actually go about processing these tasks and items as opposed to what

they are expected to be doing in producing their answers. Study Three is, in part, intended to focus on the criticism raised by Bachman et al. (1996: 129) in connection with findings of research into item difficulty, specifically, that 'very few of the content characteristics that have been identified by test developers, EFL experts, experimental research […] are actually related to item statistics'. Relying on both qualitative and, within the limits of the primarily qualitative focus of the research, quantitative analyses of the data gathered on the items, Study Three aims to explore the relationships between the item characteristics identified through content analysis and those emerging from test takers' verbal protocols, on the one hand, and the variables identified and the empirical estimates of the difficulty of the items, on the other.

It is hoped that, from a theoretical perspective, the results of this investigation, based on a triangulation of different data sources, will provide useful insights into the factors that play a part in the difficulty of matching reading tasks, the processes that are involved in responding to such reading items, thereby, at the same time, contributing to a better understanding of variables that influence performance on reading comprehension tests, in general. From a practical point of view, it is expected that the findings from the three studies will help language testers, in particular in the Hungarian context, but also in similar EFL reading assessment contexts, design and develop matching reading test tasks in such a way that scores on the test indeed reflect learners' reading abilities.

1.2 Research Questions

In accordance with the main concerns for this investigation, the dissertation aims to find answers to the following central research questions (RQ):

RQ1: *What skills, knowledge and processes are required to complete the reading items under investigation?*

RQ2: *What skills, knowledge and processes are used by the test-takers involved in the study to complete these reading items?*

RQ3: *Is it possible to observe in students' verbal reports the item characteristics identified through content analysis? In other words, do the verbal protocols provide evidence of the use of the skills, knowledge and processes predicted to be involved in responding to the items? If yes, to what extent do the two sets of data on the items agree?*

RQ4: *Is there a relationship between (any of) the item characteristic variables identified and the difficulty of the items? If yes, which item characteristics can be observed to have an impact on the difficulty of the items, and what are their specific effects? In other words, which (if any) of the item characteristics identified prove to make the items easier or more difficult to answer?*

1.3 The context of the research

The reading tasks investigated in this research were developed by the Hungarian Examinations Reform Project between 1998 and 2001, in the process of developing the new Hungarian school-leaving examination in English, which is taken by Year 12 students at the age of 18. The Project was set up in the context of overall reforms in the Hungarian educational system, which became particularly important in the light of political and economic changes initiated and implemented in and after the year of 1989, a turning point in recent Hungarian history. (For a comprehensive description of the socio-educational context of the Examination Reform, see Nikolov, 1999a.) The idea of setting up an English examinations reform project was first raised during discussions

between the Hungarian Ministry of Education, OKI (National Institute of Public Education) and the British Council Hungary in 1996 (Nagy 2000). Negotiations between the parties involved led to a four-year Agreement, signed in 1998. According to the Agreement, the British Council's responsibility included the training and the commissioning of item writers to produce items for the new school-leaving exam in English. (A detailed account of the history of the Project is presented in Nagy 2000.) The author of this thesis became a Project member in 1998 and participated in the development work related to a range of different aspects of the new exam, including, in particular, issues of the development of the reading component of the exam.

Apart from changes in the political system of the country in 1989, there were many other reasons why it became inevitable to introduce a new school-leaving examination in Hungary. These included problems with the educational system in general, indicated, among others, by the fact that, as pointed out by Nikolov (1999a: 8), 'education has lost a lot of its traditional prestige recently'. In addition, Nikolov argues, while it has been generally acknowledged that integration into the European Community requires suitable levels of foreign language knowledge on the part of Hungarian citizens, research studies investigating school-leavers' foreign language knowledge (e.g., Kádárné's [1979] IEA survey involving ten countries; a joint Hungarian-Dutch research project reported on in Noijons & Nagy 1995) have shown that, in Nikolov's (1999a: 16) words, 'few Hungarians have a good level of proficiency and […] the effectiveness of language teaching seems to be way below the desirable level', despite Hungarian learners' favourable attitudes to foreign language learning (Dörnyei et al. 1996).

From a different perspective, a study conducted by Bárány et al. (1999) to examine key stakeholders' attitudes to the Matura exam revealed that the exam had very low prestige in the eyes of virtually all groups of respondents in the study, including secondary-school teachers, headmasters, parents, school-leavers, university students and employers. According to findings of the study, some of the most apparent reasons for this unfavourable attitude to the Matura exam were related to professional aspects of the exam, to issues of reliability and validity. This is reflected in the criticism of the school-leaving exam made by secondary school teachers as follows:

- all the language skills should be covered;

- evaluation should be objective;

- standards should be higher;

- the level of difficulty ought to be consistent from year to year,

- the tasks should be more varied and lifelike (Bárány et al. 1999: 195).

For the purposes of this thesis, let us only point out the most critical aspect of the written part of the old school-leaving examination in English.

In a detailed description of the old Matura exam, Ábrahám and Jilly (1999) point out that 'more than a dozen different school-leaving tests exist in Hungary' (p. 23). However, according to their description, most test papers of the old exam followed the pattern of what was called the Basic school-leaving exam, whose written part consisted of two papers, a 90-minute Use of English test, and a translation task of the same length. Indeed, with such a strong focus on grammar and translation, the quality of the exam was highly questionable.

In the past few years, radical changes have been made to the Hungarian school-leaving examination in English. Perhaps the most significant of these is that the exam has become skills-based, that is, learners' language skills are assessed in separate tests of reading, listening, speaking, writing, and Use of English. The new exam went first live in 2005. However, as pointed out by Nikolov (2001c) with respect to the importance of quality assurance in FL teaching in Hungary, much empirical research is required to assure the quality of the new exam, including research into the effects of new task types involved in the new examination model.

1.4 The structure of the dissertation

The dissertation is structured as follows. The next chapter, Chapter 2, provides an overview of the literature. After an introduction, Chapter 2, first, discusses different views and theoretical models of reading, then presents an overview of factors that are suggested by theoretical frameworks to influence performance on language tests and, finally, examines the most important aspects of Verbal Protocol Analysis, the research methodology employed in one of the three studies (Study Two) reported on in this dissertation. Empirical studies that have investigated variables underlying tests of reading in a second or foreign language and are of significance for our investigation will be reviewed in the chapter on Content Analysis (Chapter 4).

Chapter 3 is intended to present and explore quantitative data available on the tasks and items under investigation. First, it describes crucial aspects of the process involved in the development of these reading tasks, including the item writing process and the piloting of the tasks, and then examines the difficulty of the tasks and items in the light

of empirical estimates of difficulty obtained from statistical analyses of the results of piloting. Chapters 4, 5 and 6 describe the three main studies carried out for the purpose of this research. Chapter 4 reports on Study One, which attempts to identify item characteristics likely to affect learners' scores on these reading items. Chapter 5 provides an account of Study Two, an investigation into the actual processes test takers use when responding to these tasks and items, while Chapter 6 describes Study Three, which explores the relationships among different types of information on the items obtained from the three main data sources involved in this research, that is, content analysis, think-aloud protocols, and the empirical indicators of the difficulty of the items. Although the results of the studies are discussed in the relevant chapters, the concluding chapter, Chapter 7, will provide a summary of the findings. Chapter 7 will also discuss limitations of the investigation, as well as implications for further research.

Chapter 2 Literature Review

2.1 Introduction

This study addresses as its main subject issues of the difficulty of EFL reading comprehension tests. Anyone to determine causes of the difficulty of reading tests or devise appropriate assessment procedures for such tests needs to consider the nature of reading and develop some idea of what it actually means to read texts and understand them (Alderson 2000).

Numerous attempts have been made to define both text and the nature of reading and comprehending texts. Halliday and Hasan (1976: 1) define text to have certain 'properties', specifically, certain linguistic features contributing to cohesion, 'that distinguish a text from a disconnected sequence of sentences'. A different definition describes text as language that is functional and that is used for particular purposes (Halliday 1970, 1975). Brown and Yule (1983: 6) view text as 'the verbal report of a communicative act', whilst de Beaugrande and Dressler (1981: 80) see text as 'communicative interaction', the operationalization of syntactic or grammatical structures used in real time. Widdowson characterizes reading as 'the process of getting linguistic information via print' (Widdowson 1979, cited in Alderson and Urquhart 1984: xxv), whilst Goodman (1967) describes it as 'a psycholinguistic guessing game', proposing that reading is 'a receptive language process' […] that 'starts with a linguistic surface representation encoded by a writer and ends with meaning which the reader constructs' (Goodman 1988: 12).

Clearly, reading and understanding texts, whether in the mother tongue or a foreign language, above all has to do with language and language comprehension and, to this extent, our view of reading will change in accordance with our view of language, at the extreme ends of the pendulum, from a static to a dynamic, a product to a process, a formalist to a functionalist, a structuralist to an interactive view, to refer to only a few of the opposing terms that theorists and researchers have used over the years to describe different views of and approaches to studying either language and language comprehension in general or the nature of reading.

According to recent theories, reading is a complex linguistic, cognitive, and social act and, therefore, its study, as suggested by Alderson and Urquhart (1984: xxvii), must be inter-disciplinary, involving, along with linguistics and applied linguistics, a range of other disciplines like cognitive and educational psychology, sociology and sociolinguistics, information theory, or the study of communication systems, all of which bear upon an adequate study of reading. The study of reading in a foreign language has been greatly influenced by theoretical models of, as well as research into, first language reading. While in first language reading, the complex cognitive nature of reading has been established for some time (e.g., Goodman 1967; Smith 1971; Rumelhart 1975, 1977a, 1977b; or even earlier Thorndike 1917), the traditional, predominantly 'decoding' view of second or foreign language reading has begun to change only relatively recently, dating back to the late 1970s and early 1980s (e.g., Widdowson 1978, 1983; Coady 1979; Bernhardt 1991). Carrell (1988) argues that, in contrast to reading in a native language, in the field of second or foreign language reading it has been recognized only recently that 'reading is not a passive, but an active, and in fact an interactive, process' (p. 1). The greatest impetus to change from a

decoding or bottom-up view of foreign language reading to views that emphasize 'comprehension' and the interactive nature of the reading process has come from work in the field of cognitive psychology and, in particular, the psycholinguistic model of reading, which earlier had a strong impact on views of reading in a first language (Carrell 1988).

The literature on both first and second or foreign language reading is vast, and there is a growing body of research also in the field of foreign language testing and assessment, as is reflected in the State-of-the-Art Reviews by Alderson and Banerjee (2001, 2002). Therefore, our review of the literature in this chapter must inevitably be selective, focusing on aspects of reading and its assessment that are of particular relevance to this research. The review will be presented in three parts. The first part (Section 2.2) discusses different views of reading, and examines theoretical models advanced in the literature. That is followed by an overview of factors that influence performance on language tests (Section 2.3), while the third part (Section 2.4) looks at the most important aspects of the methodology of verbal protocol analysis employed in Study Two. Note that a detailed review of empirical studies that have investigated variables underlying performance on tests of reading in a second or foreign language will be provided in the relevant chapter (Chapter 4).

2.2 Theoretical models of reading

There are various ways in which models of reading are classified in the literature. A common way is to distinguish between models that aim to describe the actual **process** of reading and those that are concerned with the result of that process, the **product**

(Alderson 2000). As the terms suggest, the former are based on a dynamic, whereas the latter on a static view of reading. That is, while process models attempt to account for the dynamic relationship between text and reader, product-oriented models typically describe reading in static terms, examine only what the reader has 'got out of' the text (Alderson and Urquhart 1984), and ignore reader variables that affect both the process of reading and the product of comprehension.

The above distinction between *product* and *process* can be related to a view of reading as *text* or as *discourse* respectively (Wallace 1992), as well as to the distinction made by Brown and Yule (1983) between **discourse-as-process** versus **text-as-product** views underlying different approaches to studying language comprehension and production in general. According to Brown and Yule (1983), a discourse-as-process view differs from a text-as-product view in that it will typically involve an investigation of 'how a recipient might come to comprehend the producer's intended message on a particular occasion, and how the requirements of the particular recipient(s), in definable circumstances, influence the organisation of the producer's discourse' (p. 24). In contrast, in a text-as-product view, while there are producers and receivers of sentences, or extended texts, the analysis concentrates on the 'product', 'the words-on-the-page' and 'does not take account of those principles which constrain the production and those which constrain the interpretation of texts' (ibid.). Brown and Yule (1983) point out that much of the analytic work undertaken in 'textlinguistics' belongs to the approach based on the latter view. As typical of such an approach, they mention Halliday and Hasan's (1976) 'cohesion' view of the relationships between sentences in a printed text, in which textuality is a property of solely the text itself.

An apparently broader view of text cohesion is reflected in the macro-structure theory developed by Kintsch and van Dijk (1978), where cohesion is seen as an instance of coherence established by the reader as the reader engages with the text. Kintsch and van Dijk (1978) examining text comprehension in terms of underlying coherence in text, draw attention to the importance of world knowledge that the reader brings to the text. They emphasise, in particular, the importance of 'schematic structures of discourse' without which, they suggest, 'we would not be able to explain why language users are able to understand a discourse as a story, or why they are able to judge whether a story or an argument is correct or not' (p. 366). In addition to their emphasis on the role of world knowledge, their approach also involves a consideration of human abilities like memory, argued to impose constraints on the comprehension process, or the reader's ability to generate appropriate inferences when necessary to maintain coherence.

In effect, Kintsch and van Dijk's (1978) model operates at the level of underlying semantic structures, which the authors characterize in terms of propositions. The propositions represent the meaning of a text and are assumed to be connected by various semantic relations (thus forming a propositional network), some of which, it is argued, 'are explicitly expressed in the surface structure of a discourse, others are inferred during the process of interpretation with the help of various kinds of context-specific or general knowledge' (p. 365). The model describes the semantic structure of a discourse at two levels, namely, at the local levels of **microstructure** and at a more global **macrostructure** level. It accounts for how a language comprehender processes the clauses and sentences of a text into a coherent semantic *text base* (microlevel), while at the same time, building up the macrostructure of the text, that is to say, developing an understanding of the text as a whole. The connection between the micro- and

macrostructure levels of the discourse is ensured by the employment of a set of specific semantic mapping rules (deletion, generalization, and construction), called **macrorules**. These macrorules, or macro-operators, working under the control of a schema (background knowledge), will reduce and organize the more detailed information in a *text base* into its gist, in other words, 'transform the propositions of a text base into a set of macropropositions that represent the gist of the text' (p. 372). In short, Kintsch and van Dijk's model, rather than trying to account for text coherence in terms of surface cohesive features, is more concerned with 'the system of mental operations that underlie the processes occurring in text comprehension' (1978: 363). In this respect, an important characteristic of the model is that, unlike some other process models (e.g., LaBerge and Samuels 1974), it assumes a multiplicity of processes occurring, as the authors claim, 'sometimes in parallel, sometimes sequentially' (Kintsch and van Dijk 1978: 364). Process approaches like that characterizing either the Kintsch and van Dijk (1978) model briefly described above or its further developed version presented in van Dijk and Kintsch (1983) are likely to have a greater potential for exploring the nature of reading than those focusing on the product of comprehension.

In support of the process approach, it is often argued that language, reading included, must be studied in process as, in Goodman's (1988: 14) words, 'like a living organism it loses its essence if it is frozen or fragmented. Its parts and systems may be examined apart from their use but only in the living process may they be understood'. The point made by Goodman appears to be in line with functional views and theories of language, which, as reflected in the work of linguists like Hymes (1972, 1974) and Halliday (1973, 1975, 1989), lay special emphasis on the user's perspective.

A central concern to functional theories is what Halliday (1973) calls '**meaning potential**'. In Halliday's (1973: 27) view, language should be seen 'as sets of options, or alternatives, in meaning that are available to the speaker-hearer' in particular social contexts and behavioural settings. For reading, such a view of language implies that, contrary to traditional views of text comprehension, meaning does not 'reside' in the text or, as Wallace (1992: 39) has put it, texts are not 'self-contained objects, the meaning of which it is the reader's job merely to recover'. Rather, meaning is created in the interaction between text and reader in the actual process of reading as the reader relies both on existing linguistic and schematic knowledge and the input provided by the text. From a different perspective, a functional view of language implies that our interpretation of texts is affected not only by psychological, cognitive, or affective factors, but also social ones (Wallace 1992). According to Wallace (1992: 43), 'our personal interpretations will never be identical with those of others […] because we have multiple social identities, any of which may be salient in our reading of a particular text'. In a similar vein, Goodman (1988) suggests that reading should be seen in its social context because, as he argues, 'the common experience, concepts, interests, views, and life styles of readers with common social and cultural backgrounds will […] be reflected by how and what people read and what they take from their reading' (Goodman 1988:13). The variation in what different people may understand from the same texts due to the social interactive nature of reading, that is, the variation in the product of comprehension, is one of the potential drawbacks of product approaches to reading.

In short, while process models are concerned with 'the entire process from the time the eye meets the page until the reader experiences the "click of comprehension"' (Samuels

and Kamil 1988: 22), product-focused models, also called **componential** models, try to 'understand reading as a set of theoretically distinct and empirically isolable constituents' (Hoover and Tunmer 1993: 4, cited in Urquhart and Weir 1998: 47). According to Urquhart and Weir (1998), componential models, as opposed to process models, 'merely describe what components are thought to be involved in the reading process, with little or no attempt to say how they interact or how the reading process actually develops in time' (Urquhart and Weir 1998: 39). As the process of reading is a silent, internal, unobservable mental behaviour, for researchers, it is easier to examine the product of reading than the processes involved. However, describing areas of skills and knowledge that might lead to comprehension is not the same as describing how the reader progressed through the text to arrive at a particular understanding. Crucial aspects of verbal protocol analysis, as a methodology that has been increasingly used to explore the process of reading, including our research, will be discussed later in this chapter.

Process models

Some models describe reading as a **linear** process which consists of a series of stages, with each stage working independently and being complete before the next stage begins (e.g., Gough 1972; LaBerge and Samuels 1974). In such models, the reader is a passive decoder of 'sequential graphic-phonemic-syntactic-semantic systems, in that order' (Alderson 2000: 17). That is, the reader begins with the visual stimulus such as printed words and proceeds, as suggested by Gough (1972: 354), 'letter by letter, word by word' to decoding the meaning of the sentence. Because the sequence of processing proceeds from recognising graphic symbols, i.e., the lowest levels of reading, to the

higher-level stages such as the decoding of meanings, linear, or serial, models are also called **bottom-up** models.

In contrast to the emphasis placed in bottom-up models on the perceptual and decoding aspects of the reading process, **top-down** approaches emphasise the importance of the reader's contribution. In top-down models, the readers, rather than being 'passive decoders', are active participants in the reading process and, in constructing the meaning of a text, they develop and use their *expectations* about the text based on *background knowledge*. The general view underlying the top-down approach is that reading is primarily **concept driven**, as opposed to being a primarily **data-driven** process as suggested by bottom-up theorists. As they emphasize the primacy of previously acquired knowledge represented by mental knowledge structures called *schemata* (Bartlett 1932; Rumelhart 1980), top-down models are also known as **schema-theoretic models**. (For related concepts, see Schank and Abelson 1977 on *scripts*, *plans*, and *goals*; Minsky 1975; and Tannen 1979 on *frames*; on applications of schema theory to stories, e.g., Mandler 1978; Mandler and Johnson 1977; Rumelhart 1975; 1977a; Stein 1982; Stein and Glen 1979; Polanyi 1982; de Beaugrande 1982; Colby 1982; Meehan 1982). To illustrate what is meant by the concept of background-knowledge-based expectations fundamental to schema-theoretic models, let us quote here an example from de Beaugrande (1982: 408), which shows how expectations work in the world of stories, more specifically, in the conversation between the Mock Turtle and Alice in Carroll's *Alice in Wonderland*:

> 'Once', said the Mock Turtle at last, with a deep sigh, 'I was a real turtle'.
> These words were followed by a very long silence […] Alice was very
> nearly getting up and saying 'Thank you Sir, for your interesting story',
> but she could not help thinking that there must be more to come.

The most frequently cited example of a top-down model is Goodman's (1967, 1971) psycholinguistic model of reading. In this model, the reader goes through a cyclical procedure of sampling the text, making predictions about what will come next, confirming (or disconfirming) predictions, and correcting them when they show inconsistencies or are disconfirmed. Decoding skills in this case are assumed to play a much smaller part in the reading process than in the case of bottom-up models. While Goodman's model is generally referred to as a top-down model, it is clearly not purely top-down, among other reasons, because it assumes that the series of cycles the reader goes through in the process of reading begins with a graphic display. Goodman (1988), despite his primarily top-down approach, accepts that first, as he says, 'the brain must recognize a graphic display […] and initiate reading' (p. 16). However, he emphasises that efficient readers do not rely much on decoding skills but rather they focus on *meaning* throughout the reading process. In Goodman's view, efficient readers, trying to get at meaning, always use strategies for reducing uncertainty, are selective about the use of the textual cues available, rely, for the most part, on prior conceptual and linguistic competence, and 'minimize dependence on visual detail' (1988: 12).

While both bottom-up and top-down approaches have had a powerful impact on reading instruction, including both first and second language reading, it is now generally accepted that 'neither the bottom-up nor the top-down approach is an adequate characterisation of the reading process' (Alderson 2000: 16). Bottom-up models are often criticized for underestimating the reader's contribution, in particular, the role of background knowledge on comprehension (Carrell 1988) and, from a processing perspective, they are criticized for their failure to account for the influence of higher level processing such as inferencing on the processing of information at the lower levels

of understanding words and sentences (Rumelhart 1977b; Stanovich 1980). Samuels and Kamil (1988: 31) have noted that especially in the case of the early bottom-up models like that of LaBerge and Samuels (1974), there was a lack of feedback loops built in the model, and so 'it was difficult to account for sentence-context effects and the role of prior knowledge of text topic as facilitating variables in word recognition and comprehension'. Top-down approaches, on the other hand, are argued to be inadequate on the grounds that they 'tend to emphasize such higher-level skills as the prediction of meaning […] at the expense of such lower-level skills as the rapid and accurate identification of lexical and grammatical forms' (Eskey 1988: 93). According to Eskey (1988), the top-down model 'is an accurate model of the skillful, fluent reader, for whom perception and decoding have become automatic, but for the less proficient, developing reader – like most second language readers – this model does not provide a true picture of the problems such readers must surmount' (p.93). However, Samuels and Kamil (1988) appear to disagree with Eskey's claim. In their view, 'while top-down models may be able to explain beginning reading, with slow rates of word recognition, they do not accurately describe skilled reading behaviours' (Samuels and Kamil 1988: 32). This clearly shows the complexity of issues involved in either modelling the reading process or assessing different aspects of such models. Discussing implications of theoretical models for ESL reading classrooms, Carrell (1988: 239) points out that some second language readers attempt to process text in a totally bottom-up fashion, which elsewhere (p. 102) she calls *text-biased* processing or *text-boundedness*, while some others process text in a totally top-down fashion, which she terms as *knowledge-biased* processing, or *schema interference*. She also notes that 'overreliance on either mode of processing to the neglect of the other mode has been found to cause reading difficulties for second language readers' (p. 239).

More recent, so-called **interactive,** models (e.g., Rumelhart 1977b; Stanovich 1980) suggest that reading comprehension is more adequately characterised as involving both processing modes (i.e., both top-down and bottom-up processes) operating interactively. They argue that efficient reading involves an interaction between **data-driven** processing and **conceptually driven** processing. A common feature of all interactive models is that, unlike top-down models, they allow for skills or various knowledge sources at various levels of processing to interact with one another in processing and interpreting the text. For example, in Rumelhart's (1977b) model, with a '*pattern synthesizer*' (the message centre) functioning as the central component of the model, the 'graphemic input' and different levels of linguistic knowledge (orthographic, lexical, syntactical and semantic knowledge) can all interact in order to arrive at 'the most probable interpretation of the graphemic input' (1977b: 588).

Grabe's (1988: 59) simplified graphic perspective on interactive processing includes the following processing levels for reading:

* graphic features
* letters
* words
* phrases
* sentences
* local cohesion
* paragraph structuring
* topic of discourse
* inferencing
* world knowledge

Eskey and Grabe (1988), supporting interactive rather than top-down models, emphasize the importance of **speed** and **automaticity** in word recognition in particular in the context of second language reading. They argue that an interactive model of reading is better able to account for the role of certain bottom-up skills that are

important to successful reading acquisition. In their view, such a model has advantages over the top-down approach in that it

> incorporates the implications of reading as an interactive process. At the same time it also incorporates notions of rapid and accurate feature recognition for letters and words, spreading activation of lexical forms, and the concept of automaticity in processing such forms – that is, a processing that does not depend on context for primary recognition of linguistic units (p. 224).

Of particular relevance to second language reading is the **interactive-compensatory** model developed by Stanovich (1980), as it is intended to take account of both skilled and unskilled reading and may thus account also for differences among proficient and less proficient readers in a second or foreign language. The basic assumption underlying the model is that, on the one hand, reading involves an array of processes and, perhaps more importantly, 'a process at any level can compensate for deficiencies at any other level' (Stanovich 1980: 36). That is, if a reader is weak in one area of knowledge or skill, for instance, at recognizing unfamiliar words or phrases in a text, they can compensate for this by strength in another area, say, by using a top-down process of guessing. If, however, the reader is unfamiliar with the topic of the text, they may decide to rely more on bottom-up processes. Grabe (1988: 61) notes that the Stanovich model 'explains many complex results of research on good and poor readers'. He supports Stanovich's claim, according to which

> compensatory-interactive models appear to be the only type of theorizing that can render certain findings in the literature non-paradoxical, such as the fact that poorer readers have been found to display larger contextual facilitation effects (Stanovich 1981: 262, cited in Grabe 1988: 61).

Text processing as modelled by 'interactionists' like Rumelhart or Stanovich represents what is called **parallel distributed processing** and, as suggested by Grabe (1988: 59), models of the type 'are often referred to as Interactive Parallel Processing models because the processing is distributed over a range of parallel systems simultaneously'.

As the term *interactive* has been used to refer to different concepts in the field of reading research, Grabe (1988) has found it important to clarify some of the meanings it may take on. He underlines that the view of reading represented by interactive models 'should not be considered as an alternate version of "reading as an interactive process"' (p. 58). As he explains, in the case of interactive models the term refers to the processing relations among various component skills in reading, whilst the interactive process of reading refers to the use of background knowledge, expectations, context, and so on; in other words, it refers to the relation of the reader to the text. In addition to the clarification of the above two meanings, he proposes for consideration in reading research a third type of interaction, which he terms '*textual interaction*'. This type of interaction refers to 'the interactive nature of the text that is being read', more specifically, 'the interaction of linguistic forms to define textual functions' (Grabe 1988: 64-65). The motivation behind Grabe's emphasis on textual interaction is that he considers 'the ability to recognize text genres and various distinct text types' as 'an important part of the reading process' (ibid.). A theoretical framework that can be used for investigating rhetorical styles and various discourse types employed in academic settings, and that is also applicable to teaching various text types is provided in Swales (1990). (Other useful work on the topic include, e.g., the model developed by Sinclair and Coulthard [1975] for the description of teacher-pupil talk in school classrooms; Winter's [1977] and Hoey's [1983, 1991] work on the clause-relational approach to text analysis; Linde and Labov [1975] on spatial networks; Nash [1985] on the language of humour; Halliday [1989] on differences between speech and writing; for a comprehensive review of Discourse Analysis and its implications for language education see, e.g., McCarthy [1991]; Hatch [1992]; and McCarthy and Carter [1994]).

As was mentioned earlier, componential descriptions try to model the reading ability rather than the reading process. Likewise, ESL/EFL reading researchers are, for various reasons, more interested in identifying and isolating components of the reading ability. They more often than not employ product-focused research methodologies, which typically involve some measure of text comprehension, often in the form of tests developed on the basis of the researchers' view of the construct of reading, and various statistical analyses are carried out to validate particular components through identifying variables that affect the results of the tests administered. Given this focus of interest in second language reading research, it is worth looking in some detail also at the componential approach. The brief overview of such models presented below draws mainly on the accounts in Alderson (2000), Grabe (2000), Urquhart and Weir (1998), and Weir et al. (2000).

Componential models

Componential models are often categorized in the literature according to the number of components identified in the models. Two-component models (e.g., Fries 1963; Venezky and Calfee 1970) generally divide reading into **decoding skills** (essentially word recognition, which may refer to recognition of graphic representation, as well as full lexical access) and **comprehension**, which is generally limited to linguistic skills, or in Fries' words, to 'a grasp of meaning in the form in which it is presented' (Fries 1963: 115, cited in Urquhart and Weir 1998: 48). The main features of two-component models are summarized in Weir et al. (2000) as follows:

> What seems to have been identified in these models are the local level decoding of lexical meanings and a global level comprehension of text with the caveat that the emphasis is in many cases laid on linguistic comprehension in these models (p. 17).

Coady (1979) and Bernhardt (1991), both describing second language reading, include three components in their models. In Coady's (1979) case these are **conceptual abilities** (which are equivalent to intellectual capacity), **process strategies** (by which Coady means both a knowledge of the system and the ability to use the knowledge, that is, language proficiency), and **background knowledge**. According to Urquhart and Weir (1998: 50), Coady's could be argued to be 'a model of comprehension and not of the reading process' as an important component, specifically, 'word recognition', is lacking in it. In Bernhardt's (1991a) model, the three variables are **language**, **literacy**, and **world knowledge**. As Bernhardt explains, 'linguistic variables entail the *seen* elements in a text, including word structure, word meaning, syntax, and morphology. Literacy variables include intrapersonal variables such as purpose for reading, intention, and preferred level of understanding, as well as goal-setting and comprehension monitoring. Knowledge entails the background information that a reader already possesses and may or may not use in order to fill in gaps in the explicit linguistic elements in a text' (1991b: 32-33, cited in Weir et al. 2000: 18).

As an alternative to the above three-component models, Carver (1982, 1983, 1984, discussed in detail in Alderson 2000) suggests that 'a simple view of reading' should include the following three variables: **word recognition skills**, **reading rate** or **reading fluency**, and **problem-solving comprehension abilities**. A further alternative is Grabe's (1991) view, in which the fluent reading process consists of six components, specifically,

- automatic recognition skills
- vocabulary and structural knowledge
- formal discourse structure knowledge
- content/world background knowledge
- synthesis and evaluation skills/strategies
- metacognitive knowledge and skills monitoring.

Among the metacognitive skills Grabe lists skills like recognising the more important information in text; adjusting reading rate; skimming; previewing; using context to resolve a misunderstanding; monitoring cognition, including recognising problems with information presented in text or an inability to understand text (Grabe 1991, cited in Alderson 2000: 13).

Elsewhere, discussing implications for reading instruction of research and model building over the past ten years, Grabe (2000: 34) claims that 'the abilities of the good reader include at least the following:

1. fluent and automatic word recognition skills, ability to recognize word parts (affixes, word stems, common letter combinations);
2. a large recognition vocabulary;
3. ability to recognize common word combinations (collocations);
4. a reasonably rapid reading rate;
5. knowledge of how the world works (and of the L2 culture);
6. ability to recognize anaphoric linkages and lexical linkages;
7. ability to recognize syntactic structures and parts of speech information automatically;
8. ability to recognize text organization and text-structure signaling;
9. ability to use reading strategies in combination as strategic readers (paraphrase, summarization, prediction, forming questions, visualizing information, skimming, scanning, monitoring comprehension, clarifying comprehension);
10. ability to concentrate on reading extended texts;
11. ability to use reading to learn new information;
12. ability to determine main ideas of a text;
13. ability to extract and use information, to synthesize information, to infer information; and
14. ability to read critically and evaluate text information.'

He notes that the above list while 'primarily drawn from L1 reading research, is also compatible with research in second language reading contexts' (ibid.).

In a comprehensive review of variables that have been proposed by theorists and/or shown by researchers to affect the nature of reading, and are relevant to the design of assessment procedures for reading, Alderson (2000: 32-84) divides variables into two main groups, those within the **reader** and those that relate to the **text** to be read. Under

reader variables he discusses background knowledge and schemata; language knowledge; knowledge of the world; cultural knowledge; skills and abilities, including general cognitive problem-solving abilities, or an ability to process information; reader purpose in reading; factors related to real-world reading versus test taking; reader motivation/interest; reader affect (the emotional state of the reader, including effects of state and trait anxiety on the reading process); stable reader characteristics like personality, sex, social class, occupation, intelligence, processing capacity in short to long-term memory, eye movements and fixations, reading speed and cognitive strategies; aspects of beginning readers and fluent readers. Among **text variables** he includes text topic and content; text type and genre; text organisation; traditional linguistic variables like sentence structure and lexis; typographical features, including layout of print on the page; aspects of the relationship between verbal and non-verbal or graphic information in text, and the medium in which the text is presented.

Several alternative views exist and over the years numerous taxonomies regarding the component skills and strategies of the reading ability have been put forward. In second language education, one of the most influential of these is Munby's (1978) taxonomy of reading 'microskills'. However, Alderson (2000: 10-11) suggests that such taxonomies should be treated with care because, as he points out,

> there is a considerable degree of controversy in the theory of reading
> over whether it is possible to identify and label separate skills of reading.
> Thus, it is unclear (a) whether separable skills exist, and (b) what such
> skills might consist of and how they might be classified (as well as
> acquired, taught and tested).

Moreover, he remarks, the origins of such taxonomies 'are more frequently in the comfort of the theorist's armchair than they are the results of empirical observation' (ibid.). Apart from the debate over implications of a **multidivisible** versus

unidimensional view of the nature of reading, that is, whether reading can be divided into identifiable subskills or it is a unitary construct, there appear to be several aspects of the reading process that are to date ill-defined and require further clarification by research. One of these is the contribution of background knowledge to reading comprehension. Definitions of background knowledge commonly involve a distinction between *general background knowledge*, or, *knowledge of the world*, and *topical* or *content knowledge*. Carrell (1988: 104) distinguishes between *formal* and *content* schemata, with the former referring to 'background knowledge of the formal, rhetorical organizational structure of the text', whilst the latter to 'background knowledge of the content area of the text'. According to critics of schema theory, the concept of a schema 'has taken on many different interpretations and it often generates as much ambiguity as it does clarity. While it is a useful metaphor for the role of background knowledge in reading, it […] is too vague to help research specify the nature and specific contribution of content knowledge' (Grabe 2000: 24). As Grabe (2000) points out, in addition to issues related to background knowledge, research findings are also ambiguous with respect to *inferencing*, *strategy use* and *metacognitive processing*. While theorists commonly agree, and there is also evidence to show, that, for instance, inferencing skills are crucial for reading comprehension, Grabe argues that the ways in which they assist comprehension 'are not entirely clear, nor is there a well-established set of inferencing skills that are readily identifiable for the improvement of comprehension, or for testing purposes' (2000: 21).

Before leaving this section, let us briefly note a major concern of reading research that applies specifically to L2 reading contexts and involves the much-debated issue of whether the ability to read transfers across languages, in other words, whether good L1

readers are also good L2 readers. From the results of relevant research Alderson (1984, 2000) concludes that there is likely to be a **language threshold** second-language readers must cross before their first-language reading abilities can transfer to reading in the second language. Clarke's (1988) "*short circuit hypothesis*" of ESL reading also suggests that limited control over the language, or limited second language proficiency may, as noted by Clarke, 'exert a powerful effect on the behaviors utilized by the readers'(1988: 119). (See also the study by Hudson 1988 on the same issue.) The implications of this, along with other relevant issues discussed at various points in this section should serve as important considerations in both reading instruction and the design and development of reading tests. Focusing on implications of research and of different views of reading for test development, Alderson (2000) points out that in developing reading tests it is important for test designers to take into account any variable that has been shown to influence either reading process or product. Testers, Alderson suggests, need to be aware that, on the one hand, 'their tests represent their view of reading' and, on the other hand, 'their view of the nature of reading, and their knowledge of the variables that can influence the reading process and the reading product, are intimately linked to the validity of their reading tests' (2000: 84).

2.3 Factors that affect performance on language tests

Performance on language tests is influenced by a range of factors and, as was briefly noted above in terms of reading tests, an understanding of the nature of these factors and, in particular, their effects on language test scores is crucial to the design and development of language tests. Considerable research in language testing has focused on construct validation, and examined the relationships between performance on language tests, or test scores and the abilities that underlie performance. The main

concern of construct validation is to demonstrate that tests measure what they are designed to measure and test scores are 'not unduly affected by factors other than the ability being tested' (Bachman 1990: 25). The concept of validity refers to the quality of test interpretation or use and involves, as described by Messick (1995: 741),

> an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment.

Bachman (1990) points out that test scores that are strongly affected by errors of measurement will not be meaningful indicators of the abilities being measured and, therefore, cannot serve as a basis for valid interpretation or use. 'A test score that is not reliable, therefore, cannot be valid' (Bachman 1990: 25). In order for our test scores to be valid and reliable, Bachman argues, we need, above all, to be able to distinguish the effects (on test scores) of the abilities we want to measure from the effects of other factors. According to Bachman, the fundamental dilemma of language testing is that

> the tools we use to observe language ability are themselves manifestations of language ability. Because of this, the way we define the language abilities we want to measure is inescapably related to the characteristics of the elicitation procedures, or test methods we use to measure these abilities. Thus, one of the most important and persistent problems in language testing is that of defining language ability in such a way that we can be sure that the test methods we use will elicit language test performance that is characteristic of language performance in non-test situations (1990: 9).

He proposes a general model for explaining performance on language tests, which is intended to provide a unified framework that could be used by both language testers and researchers for formulating hypotheses about factors that influence language test performance. In this model Bachman includes and defines four main categories of influences on language test scores. These are

- communicative language ability
- test method facets
- personal characteristics, and

- random measurement error (1990: 348).

The first category, that is, **communicative language ability**, builds on earlier theories of and extensive research into communicative competence (e.g., the Canale and Swain [1980] model, subsequently refined in Canale [1983]; the first language framework by the sociolinguist Hymes [1972], who proposed to broaden Chomsky's [1965] original concept of competence in language to incorporate communication and culture, and formulated the concept of 'communicative competence'; the work of Munby [1978] and Widdowson [1983]). While Bachman's description of communicative language ability is, in many respects, consistent with earlier work in communicative competence, his framework also extends earlier models in that it recognizes the need to explain the relationship or interaction of the components of communicative competence, an important issue earlier models tended to ignore. As Bachman states, '… the framework presented here … attempts to characterize the processes by which the various components interact with each other and with the context in which language use occurs' (1990: 81). Bachman's Communicative Language Ability (**CLA**) consists of three components: 1) *language competence*, 2) *strategic competence*, and 3) *psychophysiological mechanisms*.

1 Language competence is defined as 'a set of specific knowledge components that are utilized in communication via language' (1990:84). It includes competencies of two types: *organizational competence* and *pragmatic competence*, each of which involves further categories. Organizational competence consists of *grammatical competence* (knowledge of phonology/graphology, morphology, syntax, and vocabulary) and *textual competence* (cohesion and rhetorical organization), while pragmatic competence is seen as consisting of *illocutionary competence*, termed *functional knowledge* in Bachman and Palmer (1996), (knowledge of language

functions grouped into four macro-functions: ideational, manipulative, heuristic, and imaginative) and *sociolinguistic competence*, which is concerned with knowledge of sociolinguistic rules of appropriateness, and of cultural references and figurative language.

2 **Strategic competence** is characterized as 'the mental capacity for *implementing the components* of language competence in contextualized communicative language use' (1990: 84). As Bachman explains, [it] 'provides the means for *relating language competencies* to features of the *context* of situation in which language use takes place and to the *language user's knowledge structures* (sociocultural knowledge, real-world knowledge' (ibid.) (emphases added). (Cf. Canale and Swain's [1980: 30] strategic competence defined as '… strategies that may be called into action *to compensate for breakdowns* in communication' – emphasis added.) Strategic competence consists of three components: *assessment*, *planning*, and *execution*, with the last component, i.e., execution, drawing on the relevant psychophysiological mechanisms to implement the plan (Bachman 1990: 103).

3 **Psychophysiological mechanisms** are described as 'the neurological and psychological processes involved in the actual execution of language as a physical phenomenon' (1990: 84), which characterize the *mode* (receptive or productive) and *channel* (auditory or visual).

Before going on to discuss test method facets, it might be worth noting here that, first, in the amended model presented in Bachman and Palmer (1996), strategic competence is reconceptualized as a set of *metacognitive strategies* used in three general areas: *goal setting*, *assessment*, and *planning* (1996: 61-79). Second, the language user's *knowledge structures* has been relabelled as *topical knowledge* and, as a further characteristic of

individual language users or test takers, *affective schemata* or '*affect*' has been included as an additional component in the model (ibid.).

The second main category of influences on test scores is **test method facets**, which are described in five groups, representing different aspects of a test method. They are as follows:

1 **Characteristics of the testing environment**, which includes the test taker's familiarity with the place; equipment used; personnel; the time of testing; physical conditions; test administration.

2 **Characteristics of the test rubric** include the test organization, time allocation, and instructions. Instructions can be characterized in terms of the language (native or target), clarity of the specification of the test procedures and test tasks, and the explicitness of the criteria for correctness.

3 **Characteristics of the input** are two-fold: *format* and *nature of language*. The *format* of the input includes the channel, mode, form (language, nonlanguage, both), vehicle, and language of presentation (native, target, both), the identification of the problem, and the degree of speededness. *The nature of language* can be characterized in terms of length, propositional content (vocabulary, degree of contextualization, type of information, topic, genre), organizational characteristics, and pragmatic characteristics.

4 **Characteristics of the expected response** include all the characteristics specified for the input (see above).

5 **Relationship between input and response**: can be *reciprocal* (when interaction is involved), *nonreciprocal*, and *adaptive* (when the input is influenced by the test taker's response, as in adaptively administered multiple-choice tests). The framework in Bachman and Palmer (1996: 47-56) refines this test method facet to include, in

addition to *reactivity*, two further aspects of the relationship between input and response. These relate, on the one hand, to the *scope of the relationship*, which can be *broad* (when a task requires the test taker to process a lot of input, e.g., a 'main idea' reading item) or *narrow* (e.g., 'a short stand-alone multiple-choice grammar item', or a reading item focusing on 'a specific detail or a limited part of the reading passage') (p. 56) and, on the other hand, to the *directness of the relationship*, which can be *direct* and *indirect*, and is defined as 'the degree to which the expected response can be based primarily on information in the input, or whether the test taker […] must also rely on information in the context or in his own topical knowledge' (ibid.).

Bachman's (1990) third category of influences, that is, test takers' **personal characteristics** include, apart from attributes related to language ability, individual characteristics such as age, sex, native language, cognitive style, and affective schemata, mentioned earlier. While it is clear that language tests should be sensitive to personal characteristics, the effects of some of these factors have been more extensively investigated in language education than in language testing. For instance, although affective schemata, or emotional responses to a test task may, as is argued by Bachman and Palmer (1996: 65-66), 'influence the ways in which individuals process and attempt to complete the test task' and, more importantly, may not only facilitate but also strongly inhibit test performance, it appears that the vast majority of studies on this characteristic have been carried out by researchers in the field of language education, in particular, in second or foreign language teaching (e.g., Dörnyei et al. 1996; Dörnyei and Schmidt 2000; Nikolov 1999b).

The last category of influence is **random measurement errors**. The sources of random, or unsystematic errors include mainly unpredictable and largely temporary conditions, such as the test taker's mental alertness or emotional state, and uncontrolled differences in test method facets (Bachman 1990: 164).

Bachman (1990: 156) points out that if we want to make sure that our tests measure the language abilities we want to measure and very little else, it is important for us to understand not only the nature and extent of the effects of the factors discussed above, but also to control or minimize the effects of both test method and the interaction between test takers' individual characteristics and the test methods used in language tests. Discussing methodological issues addressed in Grotjahn's (1986) paper, he argues that construct validation studies of language tests 'must include, in addition to the quantitative analysis of test takers' responses, the qualitative analysis of test taking processes and of the test tasks themselves' (p. 270). Bachman underlines that, in his words,

> if we are to begin to understand what makes language tests authentic, and how this is related to test performance, we must also examine the *processes* or strategies utilized in test-taking as well, and this must be at the level of the *individual*, rather than of the group (1990: 335) (emphases in original).

In the next section of this chapter, we will consider a research methodology that has been increasingly used to explore the processes or strategies individuals employ in taking reading comprehension tests, namely, the methodology of verbal protocol analysis.

2.4 Verbal Protocol Analysis

Verbal protocol analysis (VPA) is an introspective technique that is widely used as a means of eliciting data about the thought processes involved in carrying out a task or

activity. It has been used extensively by researchers working in fields such as cognitive psychology, educational psychology, psychology of assessment, cognitive science, and social psychology. It is currently also used as a means for supplementing data obtained from quantitative techniques in the field of testing and assessment, increasingly playing a vital role in the validation of assessment instruments and methods (Green 1998: 2-3).

The fundamental assumption underlying introspection in general is that, on the one hand, it is possible to observe internal processes and, on the other, humans have access to their internal processes and can verbalize those processes. The theoretical framework for protocol analysis is described by Ericsson and Simon (1993). They proposed that 'cognitive processes could be described as sequences of heeded information and cognitive structures, and that verbal reports corresponded to this heeded information' (Ericsson and Simon 1993, cited in Green 1998:7). In line with this, Green (1998) argues that information that is heeded by the subject while a task is being carried out is represented in a limited capacity short-term memory and can be reported following an instruction to talk aloud or think aloud.

While protocol analysis is a type of introspection insofar as it is used to gather data by asking individuals to vocalize what is going through their minds as they are solving a problem or performing a task, in Green's (1998) view, the methodology differs from that used by early introspectionists in a number of ways. According to Green, individuals cannot directly report their own cognitive processes as is assumed by introspectionism. Therefore, as she argues, protocol analysis requires subjects to express their thoughts, but not to report the processes that produced those thoughts. The verbal protocol serves as a source of data for the researcher to infer mental processes and attended information afterwards (Green 1998).

Terminology and types of verbal protocol

The terms 'verbal protocol' or 'verbal report' are used to describe the data produced by an individual under special conditions, where the informant is asked to either talk aloud or think aloud. The 'protocol' consists of the utterances made by the individual. Verbal report data has been subcategorized in different ways. For example, Cohen (1998) describes three main categories used in second language research:

- Self-report (typically consisting of statements about general approaches to something)
- Self-observation (the subject reporting on what s/he is doing or did at the time of a particular event)
- Self-revelation (often described as think-aloud).

Gass and Mackey (2000) define such data as differing along four dimensions:

- currency (time frame, distance from event)
- form (oral, written, both)
- task type (think aloud, talk aloud, retrospection)
- support (none, full)

Shavelson et al. (1986, cited in Gass and Mackey 2000:13) outlined three types of verbal reporting (or *process tracing*): 1) Think aloud or talk aloud during a task, 2) Retrospective protocols (thinking about a previously performed task), and 3) a prompted interview, known as stimulated recall.

A most helpful description of verbal report variants is provided by Green (1998). According to the ways in which verbal protocols are gathered and the varying circumstances under which data collection is carried out, she outlines variants along three dimensions: 1) form of report, 2) temporal variations, and 3) procedural variations.

Figure 2.1 Some variations on the verbal report procedure (Green 1998:5)

	TALK ALOUD				THINK ALOUD			
Form of Report								
Temporal Variations	Concurrent		Retrospective		Concurrent		Retrospective	
Procedural Variations	Med	Non-Med	Med	Non-Med	Med	Non-Med	Med	Non-Med
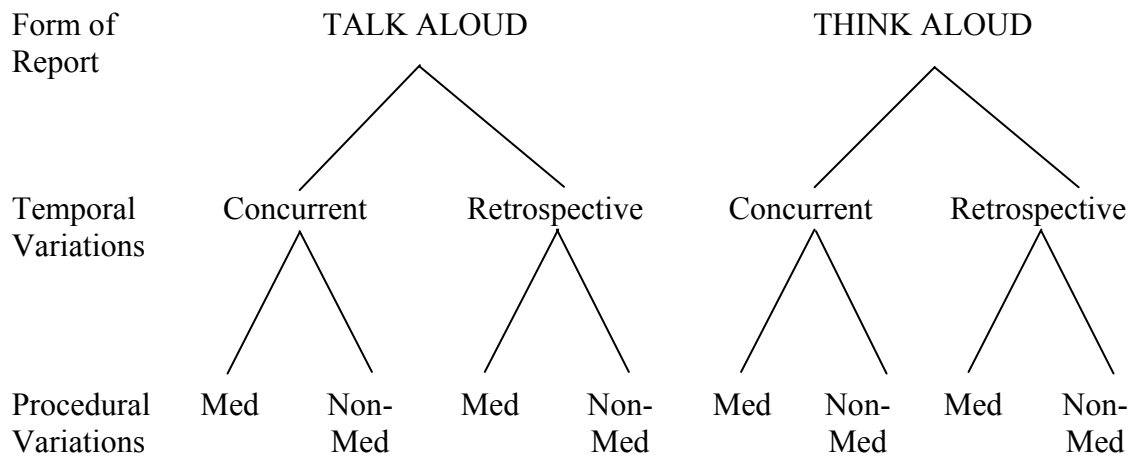
Note: Med=Mediated; Non-Med=Non-Mediated

With respect to the form of report, Green considers the distinction between talk aloud and think aloud important. In *talk aloud*, the report generated includes only information that is already encoded in verbal form. However, a task may require subjects to attend not only to verbal information but also to non-verbal visual information (e.g., about the special location of some item, graphic information, illustration, tables, diagrams), auditory information or tactile information, which first must be transformed or recoded and then verbalized. A *think aloud* report will include, apart from information already encoded in verbal form, also information that may not originally have been encoded in verbal form. To this extent, think aloud has advantages over talk aloud for tasks where subjects may attend to non-verbal information. As Green (1998:7-8) notes, the distinction between the two forms of report is quite subtle and, although it appears logical and justifiable in principle, it can be difficult to maintain in practice as subjects cannot always distinguish the two.

As can also be seen in Figure 2.1, verbal reports, whether talk aloud or think aloud, can be concurrent (simultaneous) or retrospective (subsequent). Concurrent reports are generated while the individual is working on the task, while retrospective reports are

produced after the informant has finished working on the task. It is recommended that concurrent reports are used wherever possible. One of the main reasons for this is related to potential problems resulting from the time interval between task completion and verbalisation inherent in the retrospective procedure. Even if the protocol is produced immediately after task completion, when much information can still be expected to be present in working memory, the report may be incomplete or inaccurate. Individuals might forget details of their behaviour, omit attended information and/or include information that was acquired or attended to after the task was completed. In addition, some individuals may filter information for a retrospective report in an attempt to give the impression of completeness and coherence. As a result, 'the report may be contaminated by a subject's efforts to 'tidy up' what happened, or to rationalise what occurred' (Green 1998:10). In the case of concurrent reports, these kinds of problems are less likely to arise.

Despite the advantages of concurrent verbalisations, there might be situations where the only viable option is the retrospective procedure. For example, in language testing, as Green (1998:40) points out, 'the concurrent procedure may not be usefully applied with listening and speaking'. However, when retrospective reports are used, extra care must be taken to make sure that the report generated is as complete and accurate as possible. Gass and Mackey (2000) recommend the use of stimulated recall in order to help informants to remember the thoughts that occurred to them as the task was carried out. A distinguishing feature of the procedure is that the informant is provided with some support during the recall. As Gass and Mackey (2000:17) state, 'the theoretical foundation for stimulated recall relies on an information processing approach whereby the use of and access to memory structures is enhanced, if not guaranteed, by a prompt that aids in the recall of information'. The prompt may involve a video recording or

audio recording of the event during which the participant performed a task, or it may be a piece of writing completed by the subject. Such reminders are supposed to stimulate recall of the mental processes that occurred during the original event (ibid.). (Cf. Buck 1991)

Lastly, in terms of procedural variations on verbal reports, Green (1998) distinguishes 'mediated' (prompted) from 'non-mediated' verbalisations. In non-mediated verbalisations, the informant is asked to think aloud and is prompted only when s/he pauses for a period of time. In this case, the prompts are supposed to be as non-intrusive as possible (e.g., 'Tell me what you are thinking at the moment', or a request to keep taking). In mediated verbalisations, on the other hand, the individual may be provided with more direct prompts, or may be asked questions about the task, including requests for explanations, justifications like, for example, 'Could you explain why you did that?' or 'What were you referring to just then?'.

Why is VPA used?

Protocol analysis has been used in many second language (L2) studies mainly as a means of getting at the mental processes and strategies that language learners use. Gass and Mackey (2000) list 71 studies that utilize a verbal report, with the vast majority of them (64) carried out between 1985 and 1999. The type of data in the studies listed in their table ranges from reading and writing through speaking, vocabulary and grammar to translation and L2 test-taking. The topics addressed include, amongst other areas, cognitive processes in general and specifically L2 strategy or inferencing use, L2 writing choices and processes, L2 reading and lexical use (Gass and Mackey 2000: 28-35).

One of the most useful aspects of VPA is, as Green (1998:117) notes, 'its ability to capture the dynamic nature of skilled performance. … [It] provides a wealth of information on the cognitive processes used to carry out the task, information heeded as the task is carried out, but more importantly, changes occurring in both'. This is particularly important in the light of limitations to traditional product-oriented approaches used in L2 research, including research into reading in a second or foreign language and its assessment. Alderson (1984:21-22), critiquing research in reading comprehension, points out that the data collected through the use of product approaches to investigating reading is frequently in the form of test results, most commonly, on multiple choice or cloze tests, and that 'such information provides no insight into how the reader has arrived at his interpretation, be it at the level of detail, main idea, inferred meaning, or evaluative comment'. Besides, much research in the field has been quantitative in nature, rather than qualitative. However, as Alderson argues, for adequate research in the area of foreign language reading, 'a range of information is required about individuals. … [This] would argue for in-depth qualitative studies, rather than extensive quantitative research' (ibid.).

In the field of language assessment, various aspects of language tests can be addressed using protocol analysis. Green's (1998:14-15) sampling of such questions includes, among others, the following:

- Does the test in question actually measure the set of skills it purports to measure?
- To what extent do two or more different tests that are assumed to measure the same skills actually measure the same skills?
- Are some item types more effective measures of the skill in question than others?
- Does the content of a particular item influence performance?

- Do raters heed the marking criteria in assessing performance on the task in question?
- Do rater variables influence judgements in undesirable ways?

Conducting VPA

The procedure for carrying out VPA consists of a series of distinct phases during which certain principles should be followed irrespective of the purpose of the intended application. According to the description provided by Green (1998), the first phase involves identifying a suitable task and deciding whether the task lends itself to investigation through protocol analysis. Green, however, notes that it is often the case that the 'task' is defined and it is a methodology that requires identification. As reading, listening, writing or speaking tasks are generally suited to protocol analysis (bearing in mind the need to take special care if VPA is used with listening and speaking), we rather need to specify task characteristics, because some of them may render tasks unsuitable for protocol analysis (for example, when the principal means of responding is guessing, or tasks that require Yes/No or True/False responses, or tasks which are too simple for the subjects). Besides, as pre-determined procedures and coding schemes for analysing protocols do not exist, a task analysis may help the researcher specify what is involved in responding to the task in question and consider possible ways in which the task might be completed. Information gained from task analysis may help in constructing a coding scheme later on.

The next two steps involve, first, a decision on the procedure to be used (think aloud or talk aloud, concurrent or retrospective, etc.) and, second, selecting suitable subjects.

Before data collection can begin, subjects need to be trained. This involves familiarisation with the technique itself, an explanation of what is required of them, and

the reasons for conducting the study. In addition, it is essential to provide subjects with practice in the procedure that we intend to use for the study. Preparing clear and unambiguous instructions is also essential in order to ensure that as much valid and complete data as possible is collected. If, during the session, the subject falls silent for a period of time, the use of less intrusive instructions like 'keep talking' is recommended as a prompt to continue verbalising. Unless mediation is being used, direct questions, like those mentioned earlier in this section, and requests for information by the researcher should be avoided, because they are more likely to interfere with behaviour and alter the way in which the task is approached.

With respect to its duration, the data collection session should ideally take no more than one hour, as it can be difficult for the subject to concentrate on the task, and to verbalise, for long periods of time. On the other hand, it should be taken into account that, first, as Hosenfeld (1984:232) notes, 'the length of the session will depend chiefly upon the student's ability to self-report and the interviewer's skill' and, second, the requirement itself to produce a verbal report is likely to slow down processing of the task.

Verbal reports must be gathered either by tape or on video, because these are generally considered to be the most reliable means for gathering complete information. Supplementary information (e.g., notes made by the individual as the task was carried out, or an interview with the subject after the session) may be gathered, but this is considered optional.

Once they are collected, verbal protocols are transcribed. Data transcription may be the lengthiest part of the procedure. As Green (1998) observes, it often requires a skilled audio-typist to listen to the tapes and transcribe the reports in their entirety. However, as

a great deal can be learnt through listening to the tapes, it may be very useful for the researcher to carry out some of the transcribing him or herself. In any case, the transcriber should have access to the materials that were used as the protocols were generated, because, for various reasons (e.g., background noise, or quality of equipment), it is sometimes rather difficult to understand what is being said by the subject generating the report and, in such cases, reference to those may be very helpful. An important principle to adhere to in all circumstances is that protocols should not be altered by adding or substituting words in order to achieve completeness. They should be transcribed without any modification in the original verbalisation. At the same time, it is recommended that time markers are used in the transcripts to indicate the length of pauses. Their use allows the researcher to infer, for instance, when difficulties are met by the subject. It is also desirable to check at least some transcripts against the original recordings.

In the final phase of the procedure, data are analysed. The types of analyses that are carried out will vary, depending largely on the purposes of the study. Following from the nature of the technique, standard statistical procedures cannot be directly applied to the verbal report data. Therefore, data analysis may follow two main approaches. In cases when the study aims to examine only the content of the verbal protocol in order to identify, for example, what Green (1998:100) calls 'false negatives' (responses that are wrong, but not as a result of the candidate's knowledge or skill) and 'false positives' (correct guesses), inferences may be made directly from the data without the need to quantify the data for numerical analyses. On the other hand, some studies may focus on information (e.g., frequency information) that requires quantifying the data for submission to various statistical analyses (such as, analysis of variance, or chi-square), in which case, the data need to be transformed prior to analysis. This involves

segmenting the protocols into units for analysis and then coding individual segments, using the coding categories of the encoding scheme developed by the researcher. In this way, the data gathered can be submitted to both qualitative and quantitative analyses.

Limitations to using VPA

Despite the potential for protocol analysis, there are some drawbacks to the technique, as well as some pitfalls that may be avoided, and the researcher should be aware of these. One drawback frequently referred to in the literature is that the entire process of gathering and analysing verbal protocols is very time-consuming and the procedures involved, in particular, in transcribing, coding, and analysing data may be very labour-intensive. A different type of drawback relates to the time lag between task completion and production of the verbal report. As was alluded to earlier, if there is some delay between the two, the retrieval process needs to be considered. The longer the delay, the less complete and accurate the report, because, as Gass and Mackey (2000:89), discussing findings of related studies, note, 'one's memory becomes less accurate as time passes'. This is an obvious limitation to retrospective reports in general. There are also limitations resulting from the nature of the methodology. As Alderson and Urquhart (1984:248) point out, 'thinking aloud or consciously reflecting on the process may itself distort the process'. In terms of procedural difficulties, among others, the degree of external intervention may seriously affect validity of the inferences drawn from an analysis of the verbal report. It can be very easy to disrupt and alter behaviour unintentionally, due to, for instance, inappropriate instructions. Another problem commonly referred to by critics of the technique is the issue of individual differences in the production of verbal reports, both in quality and quantity of the reports. Some individuals are more capable of verbalising their thoughts than others and produce more

detailed reports (see e.g., Alderson 1990). To date, the question seems to be unanswered 'whether sparse reports are as "valid" as more extended reports' (Green 1998:21). A final concern to mention here has to do with the choice of language for generating the verbal report in the case of L2 studies (language of the subject/L1, language of the task/the researcher/L2), which is not always straightforward, as is clear from the study by Alderson (1990).

Despite limitations to the technique, many studies in the field of L2 research, including language testing, clearly demonstrate its usefulness and its potential for providing valuable data about cognitive processes. As Green (1998:11) states, 'when the technique is used appropriately, verbal protocol analysis is a valid and useful procedure'.

2.5 Concluding remarks

In this chapter, we have reviewed theoretical models of reading, examined factors that are thought to affect performance on language tests, and discussed the most important aspects of verbal protocol analysis, the research methodology employed in our second study (Study Two in Chapter 5). In the next chapter (Chapter 3), we will look at the difficulty of the six reading tasks involved in the investigation, relying on empirical indicators of difficulty.

Chapter 3 Empirical difficulty of the tasks and items

3.1 Introduction

This chapter will present and explore quantitative data on the tasks and items obtained from statistical analyses of the results of pilot test administration of the tasks. The main aim of the chapter is to examine the difficulty of the tasks and items in the light of test and item statistics. However, before presenting and examining empirical data on the tasks, we shall first provide a brief account of crucial aspects of the process involved, and the procedures used in the development of these reading test items, focusing on the item writing process, on the one hand, and the design of piloting, on the other.

3.2 The item writing process

As was referred to earlier, the tasks selected for the purpose of this research were developed between 1998 and 2001 by the Hungarian Examinations Reform Project, as part of the process of developing the new Hungarian school-leaving examination in English. To be able to produce suitable tasks and items for the exam, above all, the then "would-be" item writers received training in item writing through attending courses and regular Item Writer workshops organised and run by experts engaged by the British Council Hungary. During the process of item production, item writers were expected to follow professional procedures developed by the Project according to international standards and practice (Alderson et al., 2000). In accordance with the aim of the Project, the item writing process was considered as part of a process leading to examination paper construction. As described in the '*Guidelines for Item Writers*' document (Alderson et al., 2000: Appendix III-3), which was drawn up to provide item writers

with both general and specific advice for producing suitable test tasks for the new exam, the first part of this process includes the following steps for item writers:

Familiarising with Specifications and Guidelines
↓
Writing items based on commissioning
↓
Small-scale pre-testing followed by revision of items
↓
Receiving feedback from pre-editing
↓
Revising items
↓
Receiving acceptance or rejection from the Editing Committee
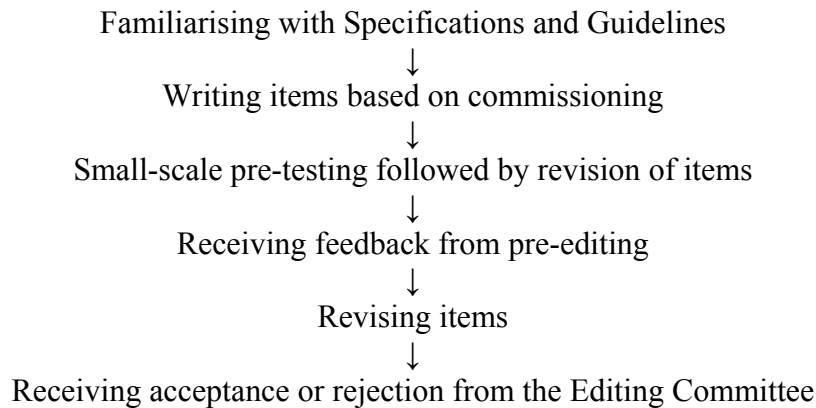
The second part of the process consists of the following:

Pre-testing items on a larger scale
↓
Revising items centrally
↓
Optional further pre-testing
↓
Building item banks
↓
Constructing examination papers
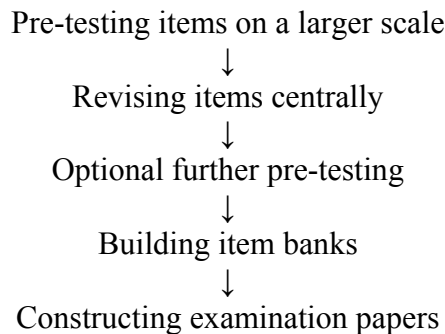
Focusing on the second part of the process described above, it should be noted that after large-scale piloting, followed by statistical analyses of the results of piloting, reading tasks that were considered by Project team members to be suitable, or required only minor changes, were centrally revised, and were then compiled, along with Use of English tasks, into a Practice Book, in order to help teachers and students prepare for the new school-leaving examination.

3.3 The piloting of the tasks

The tasks for this research were selected from three different sets of reading tasks piloted in April 1999, November 2000, and April 2001. The design of the three rounds of piloting was guided by the same principles with respect to all crucial aspects of pilot

test design, including sample size, pilot test booklet design, administration, and the analysis of the results, and the procedures followed were also the same. (For details of the procedures used, and the results of the first round of piloting, see Alderson et al., 2000).

In each round of piloting, the pilot sample involved students from different parts of the country, from different types of secondary schools (grammar, vocational, combined), and from good, average, and weak schools or classes, with the underlying aim to ensure that the tasks under development were pre-tested on relatively representative samples of the targeted test taker population. As will be clear from descriptive statistics presented in the next section of this chapter, each item piloted was responded to by at least 200 students, which enabled suitable statistics to be calculated after test administration. As in each round of piloting there were many tasks and items to be pre-tested and, therefore, a number of test booklets to be compiled and administered, it was important to develop and use a design with anchor items, that is, use a task or items common to all test booklets administered. Such a design enables test developers to compare the difficulties of tasks and items across different test booklets taken by different groups of students, by calibrating all piloted tasks and items onto a common scale of difficulty.

After pilot administration, the results were analysed, using different computer programs for various types of statistical analyses. SPSS was used for calculating descriptive statistics, Iteman for classical item analysis, and the BigSteps program for IRT-based (Item Response Theory) analyses. The results of piloting related to the six reading tasks investigated in this research are presented and discussed in the section that follows. (For the six tasks, see Appendix A.)

3.4 The results of piloting

Descriptive statistics for the six reading tasks under investigation are presented in Table 3.1 below.

Table 3.1 Descriptive statistics for the six reading tasks

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
k	10	5	6	7	6	8
n	238	238	231	258	225	254
Mean	6.794	3.592	m.d.	3.783	m.d.	4.244
S. D.	2.605	1.649	m.d.	2.539	m.d.	2.682
Alpha	0.814	0.804	0.798	0.865	0.755	0.833
Mean %	68	72	91	54	59	53
M.I-T.C.	0.620	0.750	0.712	0.743	0.671	0.679

Codes: k=Number of items; n=Number of examinees; S.D.=Standard Deviation; M.I-T.C=Mean Item-Tot Corr/mean discrimination; m.d.=missing data;

Mean % correct values in Table 3.1 suggest that, of the six tasks, the easiest is Task 3 (91%), whilst the most difficult one is Task 6 (53%). It should be noted, however, that the two tasks were piloted in different test booklets taken by different groups of students in two different rounds of piloting (Task 3 – 2001, Task 6 – 2000). As was alluded to earlier, in classical test statistics, the difficulty of an item or a test is dependent on the ability of the students taking the test. Therefore, it is worth looking also at the 'sample independent' difficulty estimates resulting from IRT analyses, which method, with an anchor task design, makes comparison across test booklets possible. Table 3.2 below shows facility values in comparison with logit values calculated for the same tasks.

Table 3.2 Facility values vs logit values for each task

Task	1	2	3	4	5	6
Facility value	68%	72%	91%	54%	59%	53%
Measure logit	-1.91	-1.42	-1.07	-0.187	1.54	1.77

In Table 3.2, Tasks 4, 5, and 6 are shown by facility values (54%, 59%, and 53%) to be of roughly the same difficulty. However, IRT-based difficulty estimates indicate a great difference in difficulties between Task 4, on the one hand, and Tasks 5 and 6, on the

other, with the logit values of the latter two tasks (1.54 and 1.77) being considerably higher than that of Task 4 (-0.187). Table 3.2 shows the superiority for analysing task (and item) difficulty over classical test theory. This is because, as noted earlier, IRT enables us to calculate item difficulty independent of the ability of the sample students being tested. Anchor items and IRT analysis allow us to compensate for the varying abilities of the sample populations tested (where the sampled students for Task 3 were evidently considerably stronger than those of the other five tasks, and the samples for Task 4 were notably weaker than those for Tasks 5 and 6). A further point to note here is the spread of difficulties across the six tasks, more specifically, the great discrepancy between the logit value of -1.91 in the case of Task 1, shown by IRT analyses to be the easiest of the six tasks, and the logit value of 1.77 in the case of Task 6, which is shown to be the most difficult one. The reliability figures for all six tasks, given the number of items involved, are quite satisfactory. As can be seen from Table 3.1, in the case of most tasks, the reliability index (alpha) is above +0.8, but even the lowest reliability, achieved by Task 5 (alpha=+0.755), is pretty close to this level. We shall now examine statistical data on the six tasks at the level of individual items. Item level results from both classical item analysis and IRT analyses are presented by task in Table 3.3 below.

Table 3.3 Item level results for each task

Task 1

Item	1	2	3	4	5	6	7	8	9	10
F.V.	50%	85%	80%	69%	89%	80%	83%	52%	54%	62%
D.I.	+.53	+.47	+.56	+.77	+.38	+.53	+.38	+.59	+.67	+.56
M	-.69	-2.79	-2.41	-1.70	-3.26	-2.41	-2.69	-.80	-.93	-1.32

Task 2

Item	1	2	3	4	5
Facility value	85%	71%	63%	78%	62%
Discrimination Index	+.35	+.63	+.66	+.57	+.79
Measure logit	-2.63	-1.42	-.52	-1.98	-.56

Task 3

Item	1	2	3	4	5	6
Facility value	93%	93%	91%	86%	90%	94%
Discrimination Index	+.23	+.24	+.27	+.34	+.27	+.20
Measure logit	-1.28	-1.23	-1.12	-.51	-.98	-1.28

Task 4

Item	1	2	3	4	5	6	7
Facility value	42%	53%	53%	42%	56%	74%	57%
Discrimination Index	+.82	+.81	+.54	+.58	+.84	+.68	+.85
Measure logit	.44	-.16	-.12	.51	-.30	-1.31	-.37

Task 5

Item	1	2	3	4	5	6
Facility value	35%	48%	79%	63%	60%	70%
Discrimination Index	+.53	+.71	+.51	+.62	+.63	+.51
Measure logit	3.06	2.26	.26	1.36	1.44	0.86

Task 6

Item	1	2	3	4	5	6	7	8
Facility value	53%	51%	41%	66%	60%	50%	49%	55%
Discrimination I.	+.70	+.73	+.71	+.68	+.78	+.71	+.78	+.77
Measure logit	1.85	1.96	2.53	.89	1.38	1.98	1.93	1.66

Discrimination indices shown in Table 3.3 are, in the case of the majority of the items, well above the generally accepted discrimination level of +0.3. The discriminative power of quite a few items can even be said to be very good (e.g., Items 1, 2, 5, and 7 in Task 4, each of which achieved a D.I. above +.8). On the other hand, relatively weak discriminations are reflected in the discrimination figures for all six items in Task 3, where, with the exception of one item, all discrimination indices are somewhat below the acceptable level of +0.3. However, if we look also at the facility values of those poorly discriminating items on Task 3, we can see that all facility values are at least very near but most are even higher than 90%, which suggests that the weak discriminations are likely to be the result of the fact the items were simply too easy for the majority of those taking the test. Difficulty estimates in Table 3.3 show that, in the case of most tasks, item difficulties vary considerably even across items on the same

task. For instance, the difficulty of the ten items in Task 1 ranges from the extremely low logit value of -3.26 to the considerably higher difficulty level reflected in the logit of -.69. There is similarly great variation in difficulty across items in Task 5, with the difficulty of the items in that task ranging between the logit values of .26 and 3.06. While it is clear that, along with the reliability of test tasks, both discrimination and difficulty are important issues to consider in test development, statistical figures alone are unlikely to provide sufficient information for developing suitable test items.

3.5 Concluding remarks

In this chapter we have looked at the empirical data available on the reading tasks under investigation. We have seen that quantitative data may reveal crucial characteristics of test items. However, before leaving this chapter, we should stress that statistical figures cannot tell us why a task or an item works the way it does, and they are not capable of revealing reasons for the identified differences in task or difficulty, either. The next three chapters (Chapters 4, 5, and 6) will explore underlying reasons for the wide range of task and item difficulties shown by the results of statistical analyses, starting with an examination of the content of the tasks and items in the chapter that follows.

Chapter 4 Study One: Content Analysis

4.1 Review of empirical studies

Content analysis has been used extensively in language testing research. It is generally considered to be essential in the development and use of language tests and plays an important part in the investigation of test validity. It has been used as a means for investigating one particular aspect of construct validity, that of "content relevance, representativeness, and technical quality" (Messick, 1995; 1996), demonstrating that a test is relevant to and covers a given area of content or ability. In Chapelle's (1999:260) terms, 'content analysis provides evidence for the hypothesized match between test items or tasks and the construct that the test is intended to measure'. Its use traditionally involves expert raters making judgements in some systematic way about the cognitive knowledge, skills and processes they believe will be required for performance on a given test (Alderson et al., 1995; Messick 1995; Chapelle 1999).

While it is generally recognized that establishing content relevance is important in the investigation of construct validity, as Bachman et al. (1996) point out, there are limitations to the use of content analysis in either test design or the process of validating interpretations of test scores. Bachman et al. (1996: 126) argue that content specification in test design 'has typically focused on the abilities to be measured, and has generally ignored characteristics of test tasks'. However, as characteristics of test methods or test tasks can affect performance on the test to a great extent, test specifications, it is argued, should include, along with information on the ability domain, also a specification of the assessment tasks, their characteristics and conditions (Bachman, 1990; Alderson et al., 1995; Bachman and Palmer, 1996; Alderson 2000;

Buck 2001). From a different perspective, Bachman et al. (1996: 126) claim that 'information about test content by itself is not a sufficient basis for score interpretations because it does not consider actual performance on tests'. Studies using test-taker introspections show clearly that, as Buck observes, 'performance on each test item by each test-taker is a unique cognitive event' (Buck, 1994: 164, cited in Brindley and Slatyer, 2002: 390), which suggests that construct validation should inevitably also involve a consideration of test taker characteristics and, more importantly, of the interaction among test content (the intended construct), task characteristics and test-taker variables.

The assumption underlying the use of content analysis is that, on the one hand, expert judges will agree on what abilities a given test or test task is testing and, on the other, content characteristics or characteristics of the test task will be related to performance on the test, i.e., to item statistics (difficulty and discrimination). However, recent research in language testing has yielded mixed results in both respects.

In a series of studies, Alderson (Alderson and Lukmani 1989; Alderson 1990a; 1990b) investigated the issue of whether experts could agree about the skills being measured by EFL reading test items, and whether there was a relationship between the level of skill (lower-, or higher-order abilities) measured by the items and the items' difficulty and discrimination. In the judgemental phase of the Alderson and Lukmani (1989) study, judges were provided with a list of skills supposedly being tested by the items and were asked to indicate against each item which skill or skills they thought the item tested. For the majority of the items, there was very little agreement on either the skills or the levels being tested (p. 263). In the empirical phase of the study, those items which the judges

had agreed upon were subjected to item analysis, and the relationships between the judged levels of the items and item difficulty and discrimination were examined. Alderson and Lukmani found 'a slight tendency for higher order skills to be more difficult', which tendency, however, was 'not marked at all' (1989: 267). On the other hand, examining the discrimination indices for each item, they observed that 'better discrimination [was] achieved by lower order items than higher order items' (ibid.). Further investigating the content of reading items, Alderson (1990b) used introspective and retrospective reports from subjects responding to multiple-choice reading test items. Results of the study provide a number of insights into the test-taking process. Firstly, test-takers 'approach the items in different ways, and different processes are involved', particularly, when they encounter unfamiliar lexical items (Alderson, 1990b: 477). Secondly, test-taker comments like that made by one of Alderson's subjects, specifically, that 'he had difficulty understanding test items, not the passages' (1990b: 468), are in support of the claim that taking reading tests involves more than solely the ability to understand texts. Thirdly, as Alderson notes, 'what appears to be being tested by an item does not always match the beliefs of judges and test constructors' (e.g., matching words between the text and the question, or eliminating implausible options), while, at the same time, 'there are .. certain skills or processes involved in and required by answering test items in a particular format that are not necessarily specified by the test writer' (1990b: 477). From his findings, Alderson concludes that 'the test-taking process (and therefore [..] at least part of the reading process) probably involves the simultaneous and variable use of different, and overlapping "skills"' (Alderson 1990b: 478).

Alderman and Holland (1981), examining differentially functioning TOEFL (Test of English as a Foreign Language) items across different native language groups, found that, for one form of the TOEFL test, EFL experts were generally able to predict on the basis of linguistic characteristics of items which items would function differentially, whereas for another TOEFL form, the predictions of the same specialists were 'too unreliable for practical application' (Alderman and Holland, 1981: 30, cited in Bachman et al., 1996).

Perkins and Linnville (1987) were interested to explore the effects of characteristics of vocabulary items on students' performance on the 40-item multiple-choice vocabulary test of the Michigan Test of English Language Proficiency. From their analyses, which included both objective measures like frequency or length of the target words and subjective ratings like how abstract a word was, they found that several features of vocabulary functioned as significant predictors. The most consistent predictors of difficulty were frequency, word length (number of syllables) and abstractness. However, results of the study also indicated that the predictive power of many variables examined varied across proficiency levels. As the authors noted, there was a tendency for different predictors to emerge at the three levels of English proficiency studied (Perkins and Linnville 1987, cited in Read 2000: 79).

Theoretical models of reading, as well as research in the field of schema theory (e.g., Goodman, 1967; Rumelhart, 1975; 1977; Johnson, 1982; Carrell and Eisterhold, 1983) suggest that 'readers will understand and recall significantly more information if texts involve a familiar topic for which they possess the relevant background knowledge' (Steffensen, 1988: 141). Focusing on this issue, Clapham (1996) examined the

interaction between language proficiency and background knowledge, in particular, the question of whether or not subject or domain specific content knowledge affected students' scores on the IELTS (International English Language Testing System) test of reading for academic purposes (EAP). For the evaluation of the content of the reading passages and test items, she used an adapted version of Bachman's (1990) Test Methods Characteristics rating scale. She reported 'quite high agreement for some of the facets', for example, "Grammar" and "Cohesion", 'although in both cases Bachman's detailed facets had been conflated into single variables', while there was little agreement among the three raters on some of the other facets (p. 151). One of the important findings of Clapham's study is that the relative importance of language proficiency and background knowledge in reading comprehension is largely dependent on the specificity of the reading passages. She observed that on 'general' passages, the level of language proficiency affected students' scores to a greater extent than did background knowledge, whereas on 'specific' passages 'background knowledge became proportionately more important' (p. 205.). However, her results also show that students do not always achieve higher scores on reading tests that are based on familiar subject matters, which is in line with findings from Alderson and Urquhart's (1985) research on the same topic. Clapham (1996) found evidence supporting the hypothesis that the effect of background knowledge varies according to the levels of students' L2 language proficiency and, more importantly, there is a threshold level below which students are unable to make full use of their background knowledge.

Bachman, Davidson, and Milanovic (1996) used content analysis in their investigation of the content comparability of six different forms of the 40-item four-option multiple-choice test of Cambridge First Certificate in English (FCE) and the relationships

between item content and item statistics. They obtained ratings by five expert judges on components of communicative language ability (CLA), on the one hand, and test method characteristics (TM), on the other, using specially designed rating instruments based on Bachman's (1990) frameworks. They achieved a high degree of consistency on test method characteristics, whereas ratings for CLA, i.e., content characteristics, showed 'much less consistency across raters than was observed for the TM ratings' (Bachman et al., 1996: 133-4). In the authors' view, based on raters' comments, the lower level of agreement among raters on the content characteristics was likely to be due, in part, to ambiguities in definitions of the characteristics and how they were understood by raters. With respect to the relationships between item characteristics (CLA and TM) and item statistics, Bachman et al. (1996) found that neither CLA nor TM ratings by themselves provided consistent predictions of either item difficulty or discrimination across the six test forms. However, when they combined TM and CLA ratings, they were able to obtain 'fairly high predictions' (p. 143). The adjusted R^2s were over .30 for five of the six regressions on discrimination and for all six on difficulty (pp. 142-143). Although most of their content characteristics (CLA) were related either to item difficulty or discrimination in at least one of the six forms examined, they were, admittedly, differentially related to these item characteristics. These results, as acknowledged by the authors, while providing 'some preliminary evidence for relationships between content ratings and item difficulty and discrimination, have raised more questions than they answered' (p. 147). For future research, Bachman et al. (1996) emphasize the importance of further revision and improvement of the rating instrument through subjective analysis of test items by test designers, item writers and content specialists.

Freedle and Kostin's (1993) research focused on predicting item difficulty for the TOEFL multiple-choice test of reading comprehension. They reviewed a large number of content analysis categories of variables from experimental studies of reading comprehension (e.g., sentence negations – Carpenter and Just, 1975; rhetorical organizers – Meyer and Freedle, 1984; Hare et al., 1989; the relative location of main idea in a passage – Kieras, 1985) which have been found to influence comprehension difficulty in non-multiple-choice testing formats, and hypothesized that many of those variables would be found to affect item difficulty in the case of multiple-choice reading items as well, and provide independent predictive information in determining item difficulty. They used a large sample of 213 items, and a subset of 98 of these items described as a non-nested sample (where each item was represented by just one item), involving three item types: main idea items, inferences and supporting ideas. Their 93 potential predictor variables were grouped into sets of '*item variables*' (e.g., item type, variables for the item's stem), '*text variables*' (e.g., vocabulary, concreteness/abstractness of text, variables for type of rhetorical organization) and '*special text-by-item interactions*' (overlap variables).

They found that 11 significant variables predicted 58% of the variance in item difficulty. Ten of these came from seven different categories of item characteristics: 1) lexical overlap between the text and the correct option, 2) sentence length, 3) paragraph length, 4) rhetorical organization, 5) the use of negations, 6) the use of referentials, and 7) passage length. The eleventh was a subject matter variable ('subject matter is social science'), which was not included in their original hypothesis as a category. Of the seven categories, there is only one, notably, the category of lexical overlap, where Freedle and Kostin identified three variables that significantly predicted difficulty,

which clearly indicates the importance of the word-matching strategy in responding to multiple-choice reading items. These variables are a) the number of words in the correct answer that overlap with words in the relevant part of the text including lexically related words for inference items, b) the percentage of words in the correct answer that overlap with words in the key text sentence including lexically related words for supporting idea items, both of which were found to make items easier, and c) the ordinal position of the earliest word on the first line of the text that overlaps with a word in the correct answer of a main idea item, which was found to be associated with harder items. The authors argue that many of the results of their TOEFL study agree with similar analyses they carried out in their earlier studies, examining SAT (Scholastic Aptitude Test) and GRE (Graduate Record Exam) reading items (Freedle and Kostin, 1991; 1992). From their findings, they conclude that 'a substantial amount of the variance [in item difficulty] can be accounted for by a relatively small number of primarily text and text-by-item predictors' (1993: 166).

In their research into the construct validity of a 45-item multiple-choice test of reading, (*Descriptive Test of Language Skills* of the Educational Testing Service), Anderson et al. (1991) examined the relationships among three types of information on the test: information on test taking strategies obtained from think-aloud protocols, information gained from an evaluation of test content, and item performance. Their content analyses were based on the test designer's categorization of each item/question as testing one of the following three aspects of reading comprehension: understanding main ideas, understanding direct statements, and drawing inferences. They used this type of analysis along with Pearson and Johnson's (1978) taxonomy of Question and Answer relationships, classifying items/questions as textually explicit, textually implicit or

scriptally implicit. Results of their chi-square analysis of the relationship between the frequencies of the test-taking strategies used by students and the question type as determined by the test developers indicated statistically significant relationships for six of the 17 reported strategies, while their chi-square statistic indicated no relationship between the strategies used and the question type as determined by the Pearson and Johnson question and answer relationships (p. 53). With respect to difficulty and discrimination, they found significant relationships between strategy use and item difficulty in the case of 9 strategies, whereas between strategy use and item discrimination in the case of only 3 of the 17 strategies. Perhaps more surprisingly, they found no relationship between item type (i.e., students' ability to understand main ideas, direct statements, and inferences) and item difficulty (p. 54). The authors conclude that 'perhaps the greatest insight gained from this investigation is that more than one source of data needs to be used in determining the success of reading comprehension test items' (p. 61).

Buck, Tatsuoka, and Kostin (1997) examined performance on a 40-item multiple-choice test of reading in a second language (one part of the reading section of TOEIC – the Test of English for International Communications), using a relatively new technique known as 'rule-space' analysis. This methodology is an application of statistical pattern-recognition techniques to diagnosing knowledge, skills, abilities, strategies, etc. that underlie test performance (Buck et al. 1997: 423). One of the most important features of the methodology is that it 'decomposes items into cognitive attributes, which represent the underlying knowledge and cognitive processing skills that the items assess; then, from the examinees' patterns of correct and incorrect responses, it infers the probability of each examinee having mastered each attribute' (p. 431). The authors drew up the

initial list of attributes or item characteristics they had hypothesized to affect performance on the basis of the research literature, linguistic theory, language teaching experience, test development practice, and self-observations of task-completion strategies, which consisted of 27 attributes. In the process of various analyses, they modified and deleted some attributes, and identified interactions (i.e., cases where two attributes occurred together). Their final list included 24 attributes – 16 prime attributes and 8 interaction attributes, with which they were able to classify 91% of test-takers into their latent knowledge states, that is, to account for the performance of the vast majority of the subjects taking the test. Despite this high classification rate, and the high correlations between the attributes and the total score on the test, because of the novelty of using the methodology for analyzing language tests, they felt it important to find some other means to cross-validate the results of their rule-space analyses, which, as they note, would ideally involve test-taker introspections (p. 444). Eventually, they used multiple regression to confirm their results. Their adjusted R^2 was .97, which indicates that, using only the attribute scores of the students who were successfully classified, they could predict most of the variance in the scores of the 40 items.

The attributes identified in the study related to textual characteristics, item characteristics, and what Freedle and Kostin (1993) termed text-by-item interactions. The list included attributes relating, among others, to basic linguistic competence, inferencing skills (e.g., 'The two items of information for the inference are scattered across the text.' / 'The ability to hold information in memory and use it to make an inference.'), background knowledge (e.g., 'It is possible to delete two distractors using background knowledge.'), the strategy of word-matching, or the fact that among the options there might be superficially plausible but wrong options (e.g., 'The most

frequently chosen incorrect option is very plausible.'). The authors found that the interaction attributes (e.g., 'The ability to understand the gist when the paragraph or segment is longer and the text is laid out in a dense continuous formatting.') were generally more difficult than the prime attributes and tended to have higher correlations with the total score, 'suggesting that they are important in explaining the performance of higher-ability test-takers' (p. 451). On the basis of the results of their study, Buck et al. (1997: 452) claim that rule-space analysis provides much useful information for construct validation of a test, as 'the attributes are the component parts of the construct, and the analysis shows exactly how much they each contribute to the total score'.

The above brief review of the literature was intended to demonstrate that researchers have employed different approaches to, as well as different methodologies in, investigating the content of reading comprehension items. Most studies have examined the relationships between item characteristics in terms of the knowledge, skills or abilities hypothesized to affect performance on the items, on the one hand, and empirical indicators of item difficulty, on the other, using a variety of statistical analyses and a range of different sets of item characteristics. Two of the studies reviewed have also included in their investigation certain aspects of the interaction between text and task (Freedle and Kostin, 1993; Buck et al., 1996), while two other, aiming to explore the processes involved in answering the items, have used verbal reports from subjects actually taking the tests (Alderson, 1990b; Anderson et al., 1991). Neither of the latter two has included task characteristics, while neither of the former two nor the rest of the studies reviewed have used test-taker introspections in their investigation of the content of the items. Lack of due attention to item characteristics involved in actual performance on the items might well contribute to the inconclusive

nature of research findings. It may be among the reasons why, as Bachman et al. (1996; 129) point out, 'very few of the content characteristics that have been identified by test developers, EFL 'experts', experimental research or theoretical models are actually related to item statistics. On the other hand, … many of the features that are most frequently used as a basis for language test design … may, in fact, not be related to actual test performance'. From a different perspective, it is also clear from the above review that the vast majority of the studies have based their investigation of predictor variables on multiple-choice reading tests, while it appears that no content analysis research at all has focused on the type of reading items investigated in this research. In the light of all this, it seems reasonable to suppose that our investigation into task and item effects may contribute to a better understanding of variables that affect performance on tests of L2 reading comprehension.

The aim of this study (Study One) was to describe the content of the tasks and items under investigation, and identify item characteristics likely to influence students' scores on these items. The main research question this study aimed to answer was formulated as follows:

RQ 1: *What skills, knowledge and processes are required to complete the reading items focused on in this research?*

4.2 Methodology and materials

4.2.1 Developing the research instrument

In order to accomplish the aim of the study, as a first step, it was important to carry out a detailed description and analysis of the tasks and items. For this purpose, a framework based on Bachman and Palmer's (1996) 'framework of language task characteristics'

was drawn up. Bachman and Palmer's model was thought to be a good starting point for the analysis because it offers the possibility to describe characteristics of the text separately from those of the tasks and items, as well as including a consideration of certain aspects of the relationship between what the authors call 'input' and 'response'. However, their framework was adapted to suit the purpose of this study. Aspects of tasks in their framework that were not relevant to this research (e.g., 'setting', or 'test rubric') were left out, while others were included in the modified framework. For instance, despite criticisms, justifiable in many respects, of readability indices, it was considered useful to check the linguistic complexity of the texts by means of using such objective ratings as well. Therefore, some readability indices were also included in the analysis. The modified framework, used to describe and analyse the tasks and items in this study, is shown in the table below.

Table 4.1 Framework for describing the tasks and items
(Based on Bachman and Palmer's 1996 framework of language task characteristics)
_____

Characteristics of the text

Text type

Language characteristics:
1) Readability of the language of texts (e.g., length of text, of passages / sentences in the text)
2) Organisational characteristics
        a) Grammatical complexity (syntax/complexity of sentences, the use of cohesive devices, discourse markers)
        b) Level of vocabulary (frequency, specificity, ambiguity)
3) Sociolinguistic characteristics (cultural references, figurative language)

Topical characteristics (familiarity, abstractness)

Characteristics of the tasks/items

Number of items on the task
Type of response / Item type
Readability of items/options (length of options)
Language characteristics of items/questions (syntax, where relevant; vocabulary)

Relationship between text and task/items

Reading (sub)skills tested
The amount of processing necessary to answer the items
The information necessary to answer an item is provided (stated explicitly) in the text
(or inference is necessary to answer the item)

As is suggested by the framework above, each task was described by the researcher in terms of 1) characteristics of the text (text type, language characteristics, and topical characteristics), 2) characteristics of the task/items (number of items, item type, language characteristics of the options), and 3) the relationship between text and task/items (skills tested, the amount of processing required, whether an inference was necessary to answer the item).

Once the description of the tasks and items was completed, an initial list of 36 item characteristics likely to affect performance on these items was drawn up, which was, however, modified and revised several times before arriving at the final set of item characteristics used for coding individual items in the second phase of the investigation. The first revision of the initial list had two main aims. Firstly, it was important to identify overlapping descriptions of item characteristics because overlaps in the descriptions would have affected the reliability of using the variables described for coding actual items. Secondly, the identification of overlapping descriptions was expected to reduce the number of item characteristics, which was thought to be rather high considering the number of items involved in the study. This revision resulted in a set of 20 item characteristics. However, even the revised list included item features described at what Buck and his colleagues (Buck et al., 1997; Buck and Tatsuoka, 1998) call the "nuts and bolts" level of item characteristics, that is, characteristics observable by the researcher, rather than a more abstract, theoretical level typically concentrating

on linguistic or cognitive aspects of item characteristics. The list included item characteristics like, for example, 'The item is ambiguous.', 'Key vocabulary in the correct answer/heading contains a lower-frequency word.', or 'Most options are syntactically possible answers to the item.' The reason for initially following an empirical approach to defining characteristics of the items is that, as Buck and Tatsuoka (1998: 125) point out, although more theoretical attributes are easier to interpret, 'the empirical attributes are far more rooted in the actual characteristics of the individual items themselves' and, therefore, as they argue, empirical researchers must begin with observable item characteristics.

However, as it was important to consider both aspects, the next step was to make inferences about the cognitive processes and abilities needed to perform the empirically described item characteristics. In some cases, it was relatively easy to define or categorise the abilities involved, in many cases, however, it was rather hard to assign one particular "ability" to a given item characteristic. There are two main reasons for this. One is that there are no clear-cut definitions in the literature of terms like '*skill*', '*knowledge*', '*ability*', *process*', '*strategy*', or the difference between '*understanding*', '*processing*', '*recognising*', '*identifying*', '*locating*', etc. The other reason is related to the fact that, in agreement with Alderson and Lukmani (1989: 264), 'a right answer may be arrived at in a variety of different ways using different processes, strategies and skills.' [..] 'One person may have difficulty with a particular word and need to infer connections across sentences, another may understand the word and, therefore, not need to infer.' (ibid.) However, if this is the case, the question arises whether, in terms of cognitive aspects, the item in question is best described as one that requires 'the ability to infer connections across sentences', or as one that requires 'vocabulary knowledge'.

The point to be made is that as soon as the researcher tries to determine characteristics of a particular item at the level of theoretical definitions, she will have to face a range of problems, including, among others, the issue of subjectivity of judgements made about the abilities needed to complete the item. Consideration of the cognitive aspects of the item characteristics entailed further modifications also in their "nuts and bolts" level descriptions.

In order to check if other experts in the field would agree with the inferences made by the researcher about the skills and processes involved in performing the item characteristics identified, it was decided to ask some members of the LTRG (Language Testing Research Group) at Lancaster University to discuss and comment on two versions of the framework available at the time. One version presented the revised list of item characteristics, the other the result of attempts at mapping the item characteristics on to abilities, processes, and strategies. These two versions are shown in the tables below, Table 4.2 and Table 4.3, respectively.

Table 4.2 Revised list of item characteristics
(version discussed by LTRG at Lancaster University, 28 Nov 2006)

1 There is lexical overlap (exact words and/or lexically related words) between the item/question and the target passage. It is possible to answer the item using a word-matching strategy.
2 The item/question has lexical overlap, apart from the correct option, with one (or two) incorrect options, as well. Selecting the correct answer requires comparing the content of two (or more) options/passages.
3 The item/question has lexical overlap with two options/passages, but the incorrect option is easy to eliminate because of the easier key words used in it.
4 The correct option/heading has no 'exact word' lexical overlap with the passage, but an incorrect option does.
5 There is lower-frequency vocabulary in the crucial information for a correct answer (in either the passage or the item/question). The key word(s) in the correct answer/option is (are) lower-frequency word(s). (Knowledge of lower-frequency vocabulary)
6 The immediate context of the necessary information (of key words) consists of a long and complex phrase or sentence. The necessary information is difficult (easy) to locate.
7 The item is ambiguous. Selecting the correct answer requires an evaluation of the information given in two or three options/passages, making an inference based on either background knowledge or the information given in the passages.
8 The 'main idea' item requires reading beyond the (three-four sentence long) passage that is in the focus of the item, and involves understanding meaning relationships across sentences of two (or more) consecutive passages/sections of the text (each of which may be gapped).

9   There are two plausible answers to the item, and the selection of the correct answer requires comparing the meaning of two options/headings with a considerable amount of semantic overlap between them.

10   There are two plausible options but it is possible to eliminate the incorrect option easily by using a word-matching strategy.

11   There are two plausible options and the elimination of the plausible but incorrect option requires reading beyond the target passage, recognizing/understanding semantic relations across sentences of two (or more) (consecutive) passages, making an inference based on information in the passages read and the correct answer.

12   The answer to the item could be an equally possible answer to (two) other items on the task, i.e., the correct answer (partially) depends on the answer(s) given to (an)other item(s) on the task.

13   The item requires the ability to process lower-frequency vocabulary, hold information in memory to make inferences based on information scattered across (larger sections of) the text.

14   The content of the passage is abstract, and the relationship between sentences of the passage is not marked with cohesive devices. (the ability to process abstract information when relationships between sentences are not explicit)

15   Matching clauses to gaps in text: Most/many options (from which to choose the correct answer) are syntactically possible answers.

16   More (than one or two) syntactically possible but incorrect options have lexical overlap with the gapped sentence.

17   Crucial information for a correct answer is included in a (single), difficult grammatical item.

18   There are two plausible answers, and the exclusion of the plausible but incorrect option requires the ability to respond to the previous (or another) item as well, or it requires the ability to distinguish fact from opinion using background knowledge.

19   The 'main idea' item requires processing larger sections of the text.

20   Crucial information for a correct response involves figurative language or an idiomatic expression.

Table 4.3 Mapping item characteristics on to skills, processes and strategies
(version discussed by LTRG at Lancaster University, 28 Nov 2006)

Skills / operations	**Item characteristic**
A The ability to use a word-matching strategy to select the correct answer	**1** The item/question has lexical overlap (exact words and/or lexically related words) with the correct option (passage) (but not with the rest of the passages. **2** There is 'exact words' lexical overlap between the correct option (heading/missing sentence or clause) and the necessary information in the (gapped) passage/ sentence.
B The ability to use other strategies than word-matching to select the correct answer	**3** The correct option (heading) has no 'exact word' lexical overlap with the passage, but an incorrect option does. **4** The correct option (heading) has lexical overlap with a passage whose understanding is in the focus of another item on the task (but not with the 'target' passage.
C The ability to compare/evaluate the content/ meaning of two or more options/passages, check the content of each passage against the item / question to select the correct answer	**5** The item/question has lexical overlap with the correct option/passage and one or more incorrect options.
D The ability to apply knowledge of easy / high-frequency vocabulary to eliminate (an) incorrect option(s)	**6** The item/question has lexical overlap with two options, but the incorrect option is easy to eliminate (e.g., because of the easier key words in it, or the item has a greater degree of lexical overlap with the correct option than with the incorrect option).
E The ability to process / knowledge of lower-frequency / 'difficult' vocabulary	**7** There is lower-frequency vocabulary (vocabulary difficult for low-level students) in

		the crucial information for a correct answer (in either the passage or the item /question / heading).
		8 The key word(s) in the correct answer is (are) a lower-frequency word(s).
		9 The passage / the immediate context of the necessary information / the context of key words has many lower-frequency words
F	The ability to understand main ideas, understand (explicit) meaning relationships within and between sentences of a short (2-3-4-sentence long) passage	**10** The item is a 'main idea' item.
G	The ability to process longer sections (two-five passages) of the text, understand (explicit) relationships across sentences of two (or more) consecutive passages of the text	**11** The (main idea) item requires reading beyond the (three-four-sentence long) target passage and requires an understanding of the relationships across sentences of two (or more) consecutive passages/sections of the text (each of which might be gapped for the task)
H	The ability to compare and interpret the meaning of two (or more) plausible options/ headings / sentences, check the meaning of each plausible option / heading / sentence against the content of the passage to eliminate (an) incorrect answer(s)	**12** One (or more) incorrect options/headings is/are very plausible. **13** An incorrect option/heading has a great amount of semantic overlap with the correct option/heading.
I	The ability to compare/evaluate the meaning of two (or more, but consecutive) options (passages) and make an inference, based on information in the passages read, or using relevant background knowledge	**14** The item/some of the options/the correct answer is/are ambiguous. The item requires an inference based on information provided in a few sentences (two or three but consecutive passages) of the text, or using background knowledge. **15** The answer to the item (partially) depends on the answer to another item.
J	The ability to synthesize scattered information, hold information in memory and use it to make an inference	**16** The item requires an inference / the elimination of (an) incorrect option(s) is based on information scattered across different sections of the text (and the sentences / clauses provided as options) **17** The answer to the item (partially) depends on the answer to another item on the task.
K	The ability to locate / recognize relevant, easily understandable information and use it to eliminate an incorrect option	**18** An incorrect option, the correct answer to another item, is very plausible, but it is easy to eliminate, because the other item involved can be easily answered by using a word-matching strategy.
L	The ability to understand abstract information (when the relationship between sentences of the passage is not explicit)	**19** The content of the passage is abstract (rather than concrete), and the relationships across the sentences of the passage are not explicit.
M	The ability to locate / recognize relevant information, process long(er), grammatically complex phrases, sentences	**20** The immediate context of the necessary information (of key words) consists of a long and complex phrase or sentence.
N	The ability to process / knowledge of difficult grammatical items/structures, understand syntactically complex sentences	**21** Crucial information for a correct answer is included in a (single), difficult grammatical item. **22** The gapped sentence is a long, syntactically complex sentence (with multiple embeddings of sub-clauses)
O	Understanding details of the content of the passage, including the meaning of the gapped sentence, and of all syntactically possible options, and compare the content of each	**23** Matching clauses to gaps in text: Most/Many options are syntactically possible answers. **24** More (than one or two) syntactically possible

	syntactically possible option against the meaning of the gapped sentence		but incorrect options have lexical overlap with the gapped sentence.
P	The ability to apply knowledge of syntax to eliminate incorrect options	25	Most incorrect options can be easily eliminated on the basis of inappropriate syntax.
Q	The ability to process / understand idiomatic expressions, topic-specific phrases, figurative language	26	Figurative language, idiomatic expressions, or topic-specific phrases are used in the information necessary for a correct response.

With the assistance of Prof. Charles Alderson, four members of the LTRG mentioned above agreed to take part in the discussion of the above two preliminary versions of the framework of item characteristics. Their discussion was tape recorded and the resulting CDs were sent to the researcher, along with a detailed written report of the notes made at the session by Prof. Alderson. Apart from providing detailed comments on each version of the framework itself, the experts involved also tried to apply the framework to categorise actual items in two of the six tasks (Tasks 1 and 2) investigated in this study. Their comments related to different aspects of the framework, including issues like the lack of clarity of certain terms used in describing the item features, the mixed use of ability categories, the lack of 'Text characteristics', to mention just a few issues raised during their discussion. Feedback from the Research Group was extremely helpful and attempts were made to take into account all their suggestions during the next revision of the framework.

Once the framework was revised in light of comments from the LTRG, it was trialled on a sample of 18 items in order to check if it was straightforward to use it for coding actual items. As in some cases, it was still difficult to assign appropriate codes/item features to particular items with confidence, some of the descriptions of item characteristics required further modifications or rewording before first coding all the items. On coding all items once, it became clear that some item characteristics were not relevant to any items involved in the study, or their occurrence was limited to one or two items. In either case, the characteristics involved were deleted, while some others

needed to be refined in order to account for the items that might have been described by the item characteristics that had been deleted for reasons of low frequencies of occurrence. The resulting set of 22 item characteristic variables, which was employed in the final coding procedure, is shown in Table 4.4 below.

Table 4.4 Final set of item characteristic variables

TEXT

Linguistic characteristics
1 Most sentences of the target section of the text are syntactically complex (as opposed to simple and compound) sentences.
2 Sentences in the target section of the text tend to be long. (The average length is above 20 words.)
3 Sentences of the target section of the text use the passive voice.
4 There is lower-frequency vocabulary (including words, phrases, idiomatic expressions) in the crucial information.

Topic
5 The content of the target section of the text is abstract rather than concrete.

ITEMS

Item type
6 The item requires locating specific information. / The ability to locate specific information.
7 The item requires identifying main ideas. / The ability to identify main ideas.
8 The item requires understanding information and recognizing structural relations within the sentence (matching-clauses-to-gaps-in-text type of items).

Language of the question/correct answer
9 Key vocabulary for understanding the meaning of the question / correct answer includes lower-frequency words or phrases.
10 Key vocabulary for understanding the meaning of the question / the correct answer includes words that might be unfamiliar to lower level students.
11 The correct answer to matching-sentences-to-text type of items is long (20 words or above) and/or involves a syntactically complex sentence.
12 There are grammatical structures in the crucial information that might be unknown to lower level students.

TASK COMPLETION / RELATIONSHIP BETWEEN TEXT AND TASK

The amount of processing required
13 A correct answer requires scanning the text.

14 A correct answer requires reading only one specific, two-five-sentence long section of the text. / The ability to understand information within one specific section of the text. (In the case of MCG type of items, within the gapped sentence.)

15 A correct answer requires reading two or more consecutive passages of the text. / The ability to understand information across two or more consecutive passages of the text. (In the case of MCG type of items, reading, apart from the gapped sentence, at least one sentence before and/or after the gapped sentence.)

Lexical overlap (between the item/question/text and the correct and incorrect options)

16 The item has lexical overlap (exact words and/or lexically related words) with the correct option, but not with the other options. / The ability to use a word-matching strategy in selecting the correct answer.

17 The item has lexical overlap with the correct option and one or more incorrect options. / The ability to ignore the lexical overlap between the incorrect options and the item, compare the meaning of two or more options, and make an inference which is more suitable.

18 The item has lexical overlap with the correct option and one or more incorrect options, but the overlap with the correct option is much stronger than with the incorrect option(s). / The ability to ignore the lexical overlap between the incorrect options and the item, compare the meaning of the options involved, and make an inference which is more suitable.

19 The item has lexical overlap with (an) incorrect option(s), but not with the correct option./ The ability to ignore the lexical overlap between the incorrect option and the item, compare the meaning of the options involved, and make an inference which is more suitable.

Elimination of superficially plausible incorrect options

20 The elimination of the (plausible but) incorrect option(s) requires comparing (two or more) options, and checking the meaning of each against the item/the relevant section of the text and making an inference based on information in the given section of the text.

21 The elimination of the (plausible but) incorrect option(s) requires an inference based on information in different sections of the text. / The ability to make an inference based on information in different sections of the text in selecting the correct answer.

Elimination of syntactically inappropriate options

22 It is possible to eliminate most incorrect options recognising their syntactic inappropriateness. / The ability to apply knowledge of syntax to eliminate incorrect options.

Using the above set of 22 variables described under eight broader categories of item characteristics, each item on each task was recoded for the characteristics supposed to affect their difficulty. Some features could be coded by clerical means (for example, counting words or clauses, checking syntactic appropriateness of a clause), whereas

others required some judgement. To minimize or, in any case, reduce the degree of subjectivity involved in judgements, some terms used in the descriptions of item characteristics had to be defined. One of these was the concept of the *plausibility of an incorrect option* (or wrong answer). An incorrect option was considered as plausible if its meaning (or content) was very similar to the meaning of the correct answer. For example, the meaning of the clause '*in 50 years' time elephants and rhino will inhabit only the echoing corridors of museums or the territory of a zoo*', given in an incorrect option, was considered to be similar to the meaning of the correct answer '*around 1,000 of our bird and animal species become extinct every year*', in particular, in the context of the sentence beginning '*the fact is that*' and, therefore, the incorrect option in question was considered to be a 'plausible' option. An incorrect option was also considered as plausible if it had support from the text even though not as much as the correct answer, which is typically the way distractors are devised. Another important concept, that of the *crucial information*, was regarded as similar to what Buck et al. (1997) called the *necessary information* and, therefore, slightly modifying their description, it was defined as 'the information in [either] the text [or the correct answer] which the reader must understand to be certain of the correct answer' (Buck et al. 1997: 437).

To enhance the accuracy of coding aspects of the language of the question and the correct answer, separate codes (v10 and v12) were included in the framework to distinguish vocabulary and grammatical structures difficult for 'lower-level students' from lower-frequency vocabulary and syntactic complexity in general. For this purpose, lower-level students (readers) were defined as being at level B1 or below according to the level descriptions provided in the Council of Europe's document (CEF) (2001). In terms of vocabulary, *basic, high-frequency vocabulary* was defined as words and

phrases required at the "Basic User" levels of A1 and A2, whilst *lower-frequency vocabulary* as words, phrases and idioms that might cause difficulty for students at level B2 on the Council of Europe scale (Council of Europe 2001: 69-71; 110; 112; 114; 224; 231; 235).

Despite efforts made to control the effects of subjectivity in coding, one major drawback of this study is that no resources were available to involve a second coder to double-code the items and check inter-coder reliability.

4.2.2 The tasks analysed

The tasks were taken from the Reading chapter of the *Into Europe: Reading and Use of English* practice book. The main principle for selection was related to 1) the type of matching involved (i.e., test method), 2) difficulty level of the task and, 3) quality of task in terms of reliability. That is, the intention was to include in the research different types of 'good' matching tasks with differing levels of difficulty. Although initially more tasks (eight) were analysed so that the data obtained from the analysis could be used to refine the framework, only those six tasks are reported on here for which calibrated IRT data and at least two students' verbal protocols were available. Of the six tasks, one involves 'matching sentences to short texts' (MST) (Task 1), two 'matching sentences to gaps in text' (MSG) (Task 2 and Task 4), another two 'matching headings to text' (MH) (Task 3 and Task 5), and one 'matching clauses to gaps in text' (MCG) (Task 6). The number of items on a task ranges from 5 to 10. The six tasks contain 42 items total.

4.3 Analysis and results

Table 4.5 below gives an indication of the difficulty level of the language of each of the six texts, as well as the options involved in each task, as rated by readability indices.

Table 4.5 Readability Indices

	Task 1	Task 2	Task 3	Task 4	Task 5	Task 6
TEXT						
Counts						
Words	356	224	377	294	502	305
Sentences	38	16	24	15	27	18
Averages						
Sentences Per Paragraph	19	2,2	3,4	1,7	3,8	4,5
Words Per Sentence	6,8	14,0	15,7	19,5	18,5	16,9
Readability						
Passive Sentences	0%	0%	20%	14%	14%	5%
Flesch Reading Ease	17,3	69,5	58,4	77,9	68,4	50,6
OPTIONS						
Words	69	96	42	183	43	101
Average length	6,2	13,7	5,2	20,3	5,3	10,0
Range in length	4-8	10-20	2-7	11-32	2-9	5-21
Passive Sentences	-	-	-	11%	-	-
Flesch Reading Ease	80,3	79,2	-	83,5	82,4	-

According to the Flesch Reading Ease scores for the texts, shown in Table 4.5 above, Text 1, with an extremely low score of 17,3, is rated as a very difficult text to read, while Text 4, with a score of 77,9, is assessed as the easiest of the six texts. These scores indicate great differences in text difficulty, in particular, between Text 1 and the rest of the texts, which is not surprising in light of the fact that the Flesch Reading Ease index bases its rating on the number of words per sentence and the number of syllables per 100 words, and it does not take into account many other aspects of the text, like, for example, its structural and rhetorical features. However, comparing the difficulty of Text 1 and the rest of the texts in terms of only the number of words per sentence and/or

the number of syllables may be misleading, since the great difference (6,8 words per sentence in the case of Text 1, and over 14 words in each of the other texts) is likely to result from the fact that Text 1 consists of sixteen 'small ads', which are typically short texts and, more importantly, use fragmentary language consisting of, for the most part, incomplete sentences. This, however, does not necessarily make Text 1 'very difficult' to read, as its low Flesch Reading Ease score would indicate, and shows the weakness of the formula itself. For purposes of comparison, Table 4.5 gives more useful information about the length of the sentences across the six texts, showing that Texts 4 and 5 tend to use longer sentences than the others. With respect to the options, from which to choose the correct answers to the items, the longest options are also used in Task 4. The task by task analysis below provides more detailed information about the characteristics of each individual text and task.

4.3.1 Task level analysis

<u>Task 1</u> (Julie wants …) (See Appendix A)

CHARACTERISTICS OF THE TEXT

Text type: Classified advertisements taken from newspapers and magazines

Language characteristics

1   Organisational characteristics

a) Grammatical complexity:

Apart from a small number of syntactically complete sentences, the organization of information in the text is phrasal. Some phrases used are very short and simple (e.g., '*Wedding and portrait photography.', 'Unique gifts from £35'.*) In some cases, information is provided in the form of a simple list of nouns (e.g., '*People, children, pets, houses and cars*'.), while a few advertisements contain relatively long and

complex phrases (e.g., '*quality hand waxed furniture traditionally constructed in antique style, from new wood*').

b) Level of vocabulary:

Roughly one third of the vocabulary consists of basic, high frequency words (e.g*., ideal, persons, shoes, water, dentist, photography, television, video, long hair, perfect place, live music, romantic meal, antique clocks, Christmas present*). Words and phrases in the text that might be 'difficult', in particular, for lower level students (students at level B1 and below) include, for example, *measure, equipped, crafted, waxed, accredited, affordable, delivery, installation, professional consultants, competitive rates*, and some specialised words related to particular topics (e.g., *manuscript, perm, bore hole drilling*).

2   Sociolinguistic characteristics

The text contains a number of references to social and cultural values. (e.g*., the Bob Moffatt Jazz Quartet, a 50ft Hinckley yacht, Victorian style lamp posts*). Besides, the names and addresses of the advertising companies all involve culture specific proper names and geographical names (e.g., *Catteral & Wood Ltd; La Tama, Ainsworth Village; Calf Haugh Farmhouse, Pateley, North Yorks; Malcolm Eckton Studio, 18 Berry Lane, Longridge*).

No figurative language is used in the text.

Topical characteristics

Although the text covers many different topics, they are all concrete topics and most are related to everyday life.

CHARACTERISTICS OF THE TASK

No of items: 10

Item type: matching sentences to short texts / advertisements (MST)

Language of items/questions:

a) Grammatical complexity: Each question consists of a short, simple sentence, with all ten items/questions using the same simple sentence structure, the basic SVO (+Adv) order. Besides, the same verb "*want*" is used in the Present tense in all ten questions.

b) Vocabulary: Most items/questions involve high frequency words and phrases (e.g., *sandals, eat out, doctor, band, hairdo, exotic trip, pictures for business*).

<u>Note</u>: The words '*valuable*' in Item 1, and *'(home) entertainment'*, in Item 8 may be difficult for some low-level students. The word '*glasses*' in Item 9 involves a degree of semantic ambiguity.

Reading (sub)skills tested: ability to find specific information

<u>Task 2</u> (Giant Pandas) (See Appendix A)

CHARACTERISTICS OF THE TEXT

Text type: A descriptive text from a magazine

Language characteristics

1   Organisational characteristics

a) Grammatical complexity:

Most sentences are either simple (8) or compound (4), and there are 5 complex sentences. Most sentences are relatively short (below 14 words in length). The two longest sentences contain 18 and 19 words. The most frequently used sentence connector is the copulative "*and*", "*but*" appears twice in the text, and two of the complex sentences use the subordinating conjunctions "*once*" and "*when*". All sentences are in the Present tense, employing mainly high frequency verbs like "*be*" (7x), "*live*" (3x), "*eat*" (3x), "*have*" (2x), "*like*", "*grow*", "*search*", "*learn*", "*find*",

"*climb*", "*look (alike)*", "*takes (time)*", "*stay (away)*". Most grammatical structures used are 'basic', very easy structures.

b) Level of vocabulary:

The vast majority of the vocabulary consists of simple, easy, high frequency words. The more difficult words and phrases are, for the most part, related to the specific topic of the text (e.g., *stems, twigs, shoots of young bamboo plants, mammals, males, females, scent glands, mate, area, miles, diameter*).

2   Sociolinguistic characteristics

The text contains no social or cultural references that are likely to affect comprehension. No figurative language is used in the text.

Topical characteristics

The text provides basic, factual information about giant pandas. The topic is concrete.

CHARACTERISTICS OF THE TASK

No of items: 5

Item type: matching sentences to gaps in text (MSG)

Note: The two longest options are Option C (16 words), which is the only distractor included, and Option G (20 words), the correct answer to Item 3.

Language of options

a) Grammatical complexity: Of the seven options, four involve syntactically simple sentences, two are compounds and one is a complex sentence. The grammatical structures, similarly to most of those that appear in the text, are very easy.

Note: The only complex sentence among the options, involving a concessive clause introduced by the subordinating conjunction "*although*" is Option G, the correct answer to Item 3.

b) Vocabulary: Most options involve simple, easy, high frequency words. The most difficult vocabulary items appear to be "*share*", "*territory*", "*get close to*", "*cave*" and "*shelter*", with the first three appearing in Option G, the correct answer to Item 3, the other two in Option C, the distractor.

Reading (sub)skills tested: ability to understand main ideas

<u>Task 3</u> (Gorillas in …) (See Appendix A)

CHARACTERISTICS OF THE TEXT

Text type: A descriptive-narrative text from a newspaper

Language characteristics

1   Organisational characteristics

a) Grammatical complexity:

Most sentences of the text (14 out of 24) are syntactically simple, the rest are complex sentences, typically involving one sub-clause each. The syntactically simple sentences, in some cases, use fairly long, verbal phrases (including infinitive, gerundial and participial phrases as well), often put either as an adverbial modifier or the subject of the sentence in sentence initial position. The most frequent type of sub-clause is the relative clause. The two main tenses used are the Present tense and Past Simple. A relatively high number of sentences use the verb "to be". (It appears 10 times in the text as the main verb.) The vast majority of the verbs are stative verbs. Passive structures are used in a few cases only and, with the exception of one "must", there are no modals in the text. The relationships between sentences/parts of the text are, for the most part, explicitly marked with adverbials, co-ordinating and subordinating conjunctions and other discourse markers like *even so,* or *to such an end*.

b) Level of vocabulary:

Some of the vocabulary consists of basic, high frequency words and phrases (e.g., *face, home, forest, family, mountain, survival, daily lives, feelings, food, wide, clear, huge, magical, wild animals, modern world, local people, the biggest dangers, the heart of Africa, feel/felt, sleep, played, watched, brought*). However, the ratio of more difficult, less-frequently used words is at least as high as that of the simple and easy ones (e.g., *hacking through, stared out, gazed, munched, designated, split, ploughed back into, tinged with mischief, transfixed, mist-shrouded, tremendous, primeval, innocence, intruder, troop, sustenance, the edge of extinction, revenue sharing*).

2   Sociolinguistic characteristics

The text contains no cultural references that are likely to make comprehension more demanding for Hungarian students. There are, however, a few instances of figurative language use (e.g., the survival of the gorillas *hangs by a thread*).

Topical characteristics

The topic of the first four paragraphs is concrete, while the information in the last two paragraphs is abstract rather than concrete.

CHARACTERISTICS OF THE TASK

No of items: 6

Item type: matching headings to paragraphs of a text (MHT)

Language of options/headings

a) Grammatical complexity: Of the eight options/headings, four take the form of a noun phrase; a noun clause (introduced by "What" or "How") is used in three options; and one involves a gerundial phrase. Most headings are grammatically fairly simple. Two of them use the passive voice (Option A, the answer to Item 2, and Option G, the answer to Item 6.)

b) Vocabulary: Most headings contain simple and easy words (e.g., *gorillas, population, reaction, group, meet, protect, did*). The words *leader*, *author*, *experience* and *appreciation* might be unknown to lower level students.

Reading (sub)skills tested: ability to understand main ideas

<u>**Task 4**</u> (Being wet …) (See Appendix A)

CHARACTERISTICS OF THE TEXT

Text type: A narrative of personal experience taken from a teenage magazine

Language characteristics

1   Organisational characteristics

a) Grammatical complexity:

Of the 15 sentences of the text, 8 are 'composite' sentences, containing a mixture of compound and complex sentences, with at least two main clauses in each and, besides, there are 3 further sentences that are either complex or compound. All this results in a length of around 20 words or above in most sentences of the text, with the longest containing 30 words. Apart from a few instances of ellipsis, the relations within and between sentences as well as across larger sections of the text are made explicit, for the most part, with fairly simple cohesive devices. The most frequently used co-ordinating conjunctions are "*and*" and "*but*", while in the case of complex sentences, the vast majority of sub-clauses are either adverbial clauses (of time) using conjunctions like "*when*" (3x), "*as*", "*until*", "*while*" or noun clauses functioning as the object of the verb and using "*that*". Two further conjunctions used are "*so*" and "*because*". Indirect Speech appears in some (4) sentences, two of which present indirect questions introduced by "*if*". The verb tense in the vast majority of sentences is the Past Simple tense, with Past Perfect appearing, as a result of 'tense-shift', in some of the sub-clauses

of the sentences involving Indirect Speech. Two reported statements are in the passive.

The majority of the verbs in the text are basic, high frequency verbs, like *wanted, was/were, took, got (in)to, had, asked, said, told, looked at, watched, sat, (couldn't) believe, came, rang (up/back)*, reporting main events of the narrative. Verbal phrases in some sections of the text may increase the difficulty of those sections.

b) Level of vocabulary:

Most of the vocabulary consists of easy, high frequency words. However, the text also contains some more difficult words and phrases like *expect, ban, shriek, cover, a clap of thunder, pull away, avoid, treat, spoil, hold an investigation, guard, off duty*.

2   Sociolinguistic characteristics

The only cultural reference in the text is made to *South West Trains*, the company involved in the story. Some instances of figurative language use, involving phrasal verbs, lexical phrases or idioms, like *pick on sb, be fed up, end with a bang, give them a send-off*, may present some difficulty, in particular, to lower level students.

Topical characteristics

The topic of the text is a concrete, everyday topic *(travelling by train)*.

CHARACTERISTICS OF THE TASK

No of items: 7

Item type: matching sentences to gaps in text (MSG)

Language of option:

a) Grammatical complexity: Among the 9 options, there is only one simple sentence, the rest include two composite, three complex and three compound sentences. For the most part, the options use the same type of grammatical structures as the sentences in the text. Their length shows a wide range.

Note: Grammatically, the two most difficult sentences are Options A and B, the answers to Items 4 and 6, both using a hypothetical conditional, with Option B using it in an indirect question. Besides, Option B uses a further complex structure, which involves a modal auxiliary (*might*) to express past possibility in a passive clause. These are also the two longest options. Options C and D, the answers to Items 2 and 3, use a passive structure, the rest of the options are grammatically easier. The shortest and grammatically easiest option is Option F, the correct answer to Item 1, which involves a simple sentence using the Past Simple tense.

b) Vocabulary: Similarly to the text, most of the vocabulary used in the options can be said to be fairly easy. However, there are also some lower-frequency words and phrases like *hardly, wreck, drenched, tramped around, wandered (back) and eventually,* with the latter two appearing in Option F, the correct answer to Item 1, which is grammatically the easiest option.

Reading (sub)skills tested: ability to understand main ideas

Task 5 (Caught out in the rain) (See Appendix A)

CHARACTERISTICS OF THE TEXT

Text type: A mainly narrative text taken from a newspaper

Language characteristics

1   Organisational characteristics

a) Grammatical complexity:

Most sentences of the text are syntactically complex sentences. One of the complex sentences (in Paragraph 2) consists of 49 words and involves six sub-clauses (4 object clauses, a relative clause and an adverbial clause of reason). Many sentences, including most syntactically simple sentences, use relatively long verbal phrases (participial,

gerundial, as well as infinitive phrases) as postmodifiers, or placed in front position to introduce the sentence, or occasionally both. The three main verb tenses in the text are the Past Simple, Present Simple and Past Continuous. Passive structures and Modals (*can, could, may*) are used only in a few sentences.

b) Level of vocabulary:

The text uses many low frequency words and phrases (e.g., *downpour, attitude, salvation, credit(ing), prescience, rural, urban, lodged, transferred, outweighed, confirm, retrace, detour, accrues to, glancing through, leaned over, rat run, eagle eye, steely expression, adjoining desk, public domain, corporate image, atavistic approach, vast sewer reconstruction system*).

2   Sociolinguistic characteristics

There are some instances of figurative language use, including idiomatic expressions and phrasal verbs (*seemed at hand, be on one's mind, get off [my land], turned out [to be], caught out, be up to*).

Topical characteristics

Two topics are covered in the text. One (the story of getting to an appointment) in the first five of the seven paragraphs of the text, and another one (the use of commercial office buildings) in the last two paragraphs. The content is concrete in the first five paragraphs, and abstract in the last two.

CHARACTERISTICS OF THE TASK

No of items: 6

Item type: matching headings to paragraphs of a text

Language of options/headings

a) Grammatical complexity: Grammatically, all options/headings are fairly simple. Of the eight options, six contain noun phrases, two of which are somewhat longer (Options

C and D) than the others. One option consists of an easily understandable saying ("*An Englishman's home is his castle*").

b) Vocabulary: The options are more difficult in terms of vocabulary than grammar. Many of them use low frequency words. For example, Options B ("*An unexpected narrow escape*"), F ("*A sudden obstacle*"), C ("Two *approaches to public use…*") and G ("*Possible short-cut*").

Reading (sub)skills tested: ability to understand main ideas

<u>**Task 6**</u> (Animals under threat) (See Appendix A)

CHARACTERISTICS OF THE TEXT

Text type : A mainly expository text from a magazine

Language characteristics

1   Organisational characteristics

a) Grammatical complexity:

The text contains many complex sentences. Most sentences are not particularly long: even the two longest contain only 28 words each. However, the vast majority involve relatively difficult and, occasionally, rather complex grammatical structures. These include, among others, many verbal phrases, used as adverbial modifiers, or as pre- or post-modifiers to a noun, or, as part of a longer prepositional phrase or an elliptical defining relative clause (e.g., "*campaigning conservationist", "Nature's way of ensuring the survival of…", "those best suited to …*"), an adverbial clause of purpose introduced by "*so that*", a conditional sentence using the relatively difficult conjunction "*unless*", the Split Infinitive ("*help to artificially save*"), occasionally rather long, prepositional phrases used as free modifiers to introduce the sentence (e.g., "*For generations of children learning to read, ..*", "*In one clutch of eggs from a giant*

tortoise, for example, ..."), or as post modifiers ("*Sophisticated techniques, from test tube fertilisation to embryo freezing, can...*"). As a result of employing sometimes rather complex phrases, most syntactically simple sentences in the text appear to be relatively long for the type and, more importantly, probably become no less difficult to process than some complex sentences of the text. The two main verb tenses are the Present Simple and the Future tense with "*will*". There are also two conditional sentences, three or four instances of the Present Perfect, while the Present Continuous is used in one sentence. Passives occur twice in the text. Modals include *could, would* and *should,* each appearing once in the text, while *can* and *be able to* are used twice each.

b) Level of vocabulary

The text uses a relatively high number of lower-frequency vocabulary items (e.g., *gloomy, exaggerated, sophisticated, alarmist, identical, continual, echoing, starve, transmit, adapt, pacing [up and down], gape at, inhabit, surroundings, characteristics, drought, process, threat, territory, failure, flick*), including many topic-specific words (see those below).

2   Sociolinguistic characteristics

Cultural references in the text include references to two naturalists (Charles Darwin, David Attenborough), a protection group (Elefriends), Darwin's book, *The Origin of Species by Means of Natural Selection* and a children's book, Rudyard Kipling's *Jungle Stories*. No figurative language is used.

Topical characteristics

The topic of the text is the problem of the extinction of endangered species of animals. The content covered is mainly abstract. Although some of the topic-related words in the text are high frequency, easy words (e.g., *bird, elephant(s), tigers, rhino, animals, creatures, zoo(s), zoologist, wildlife, survival, save*), there are many relatively difficult,

less frequently used words and phrases as well (e.g., *preserve, develop, evolved, evolution, protection, extinct(ion), conservationist, naturalist, natural habitats, endangered species, genetic make-up, clutch of eggs, hatchlings, offspring, test tube fertilisation*).

CHARACTERISTICS OF THE TASK

No of items: 8

Item type: matching clauses to gaps in text

Language of options

a) Grammatical complexity: The grammatical structures used in the options show the same wide range as in the input text. Option D uses the Present Perfect tense in the passive, which is likely to be unknown to lower level students, Option H consists of a 14-word long infinitive phrase, Option H uses the Present Perfect tense along with the Disjunctive "*either ... or*".

b) Vocabulary: It is roughly of the same difficulty level as the vocabulary in the input text. In terms of vocabulary, the two most difficult options seem to be Options J and H, the correct answers to Items 8 and 5. The two easiest are Options D and G.

Reading (sub)skills tested: ability to understand text structure

4.3.2 Item level analysis and results

In addition to a descriptive, item-by-item analysis of what is supposed to be involved in answering the items, codes of the item characteristics assigned to each item on the basis of the framework of 22 item characteristic variables described earlier in this chapter are shown next to each item.

TASK 1 (Julie wants …)

Text type: Classified advertisements taken from newspapers and magazines

Item type: Matching sentences to short texts

Reading (sub)skills tested: ability to find specific information

Item 1   Expected Variables (EV): *v6, v10, v13, v17*

'Jack wants something old and valuable.'

The word '*valuable*' in the item/question is likely to be unknown to lower-level students. The item has lexical overlap with two options, the correct option, and an incorrect one. The correct option talks about '*antique clocks*', while the incorrect option offers '*antique style*' '*furniture*' '*from new wood*'. The selection of the correct answer requires comparing the content of two options, and making an inference which is more suitable.

Item 2   EV: *v6, v13, v16*

'Jill wants a new pair of sandals.'

The item has lexical overlap with the correct option, but not with the other options. It requires matching the words '*shoes*' and/or '*feet*' used in the correct answer with the word '*sandals*' in the item.

Item 3   EV: *v6, v13, v16*

'*Angela wants to eat out with her boyfriend.*'

The item has lexical overlap with the correct option, but not with the other options. It is possible to give a correct answer by identifying the match between the words '*eat out*' in the question, and '*romantic meal*' in the correct answer.

Item 4   EV: *v6, v13, v16*

'Charles wants to go on an exotic trip.'

The item has lexical overlap with the correct option, but not with the other options. A correct answer can be given by identifying the match between the words '*exotic trip*' in the question and '*coast*', '*Turkey*', '*yacht*' in the correct answer. Although the key words are slightly more difficult than in the case of some other items in the task, they are likely to be familiar even to most low level students.

Item 5   EV: *v6, v13, v16*

'*Cathy has a toothache and wants a doctor.*'

The item has lexical overlap with the correct option, but not with the other options. The correct answer can be selected by identifying the match between the words '*toothache*' and '*doctor*' in the question and the word '*dentist*' in the correct answer.

Item 6   EV: *v6, v13, v16*

('*Richard wants a band for his party.*')

The item has lexical overlap with the correct option, but not with the other options. It requires students to identify the match between the words '*band*' and '*party*' in the question, and '*dinner jazz*', '*live music*' in the correct answer.

Item 7   EV: *v6, v13, v16*

'*Jane wants a new hairdo.*'

The item has lexical overlap with the correct option, but not with the other options. It is possible to give a correct answer by identifying the match between the words '*hairdo*' in the question and '*long hair*' in the correct answer.

Item 8   EV: *v6, v10, v13, v18*

'*Peter wants home entertainment.*'

The item has some degree of lexical overlap with three options, the correct option (offering '*television and video equipment*') and two incorrect options (using words like '*jazz*', '*music*', '*romantic meal*'), but the overlap with the correct option appears to be

stronger than with the other two options (as television and video are more typically tied to 'home'). The selection of the correct answer requires comparing the content of the three options and making an inference. Key vocabulary for understanding the meaning of the question includes a word ('*entertainment*') that might be unfamiliar to lower-level students.

Item 9   EV: *v6, v10, v13, v16*

'*Jessica wants new glasses.*'

The item has lexical overlap with the correct option, but not with the other options. The correct answer can be selected by identifying the match between the words '*glasses*' in the question and '*spectacles*' or '*eyewear*' in the correct answer. One of the key words ('*spectacles*'), which is used in the body of the advertisement, might be unfamiliar to lower level students, whilst the other, the easier one ('*eyewear*') is used in the heading of the advertisement, and therefore it is easier to overlook.

Item 10   EV: *v6, v10, v13, v17*

'*Roger wants pictures for his business.*'

The item has lexical overlap with three options, the correct option, using the word '*photography*', and two incorrect ones, in which the words '*photography*', '*portrait*', and '*portraits*' are used. Selecting the correct answer requires comparing the content of the three options and making an inference which is more suitable. Key vocabulary for understanding the meaning of the correct answer includes words that might be unfamiliar to lower level students ('*advertising*' and/or '*commercial*' [photography]).

TASK 2 (*Giant Pandas*)

Text type: A descriptive text from a magazine, providing factual information

Item type: Matching sentences to gaps in text

Reading (sub)skills tested: ability to understand main ideas

Item 1   EV: *v7, v14, v16*

It is possible to answer the item correctly by reading only the three-sentence long section of the text in which the gap for the item is located. The item requires understanding the main idea in the given section of the text and recognizing the relationship between the topic of the section and the meaning of the correct answer (that is, the meaning of the sentence that fits the gap in the section). There is considerable lexical overlap between the given section and the correct answer. The overlap includes words like '*eat*', '*plants*', '*bamboo*' used in the text, and '*plant eaters*' and '*a plant ..*' '*bamboo*' in the correct answer, that is, the sentence that fits the gap in the section.

Item 2   EV: *v7, v14, v16*

The item requires reading the two-sentence long section of the text in which the gap for the item is located, understanding the main idea in the given section, and recognizing the match between the topic of the section and the content of the correct answer. Crucial information in both the given section of the text and the correct answer (the missing sentence) is provided with simple, easily understandable phrases and expressions related to height and weight, which makes it easy to answer the item (the phrases '*1.5 m tall*' and '*as tall as 1.7 m*' are used in the text, while '*weigh as much as 117 kg*' is used in the correct option).

Item 3   EV: *v7, v10, v11, v14, v16*

It is possible to answer the item correctly by reading only the two-sentence long section of the text in which the item is located. Selection of the correct answer requires understanding the main idea of the section and recognizing the relationship between the meaning of the section and the meaning of the correct answer. There is lexical overlap between the text and the correct answer, with the words '*area*', '*miles*' and '*diameter*'

used in the given section of the text, and the word '*territory*' appearing in the correct answer. Key vocabulary for understanding the meaning of the correct answer (the word '*territory*') might be unfamiliar to lower level students. Of the 7 options, the correct answer is the longest (20 words) and only option involving a complex sentence.

Item 4   EV: *v7, v14, v18*

It is possible to answer the item correctly by reading only the three-sentence long section of the text in which the gap for the item is located. It requires understanding the main idea in the given section, and identifying the relationship between the meaning of the section and the meaning of the sentence that fits the gap in the section. The particular section has lexical overlap with the correct option and an incorrect one, but the overlap with the correct option is much greater than with the incorrect option. The overlap with the correct answer includes '*A new-born panda*' in the first sentence of the section and '*Pandas are born*' at the beginning of the sentence that is considered as the correct answer to the item.

Item 5   EV: *v7, v15, v16*

The section in which this item is located is the last section of the text. It consists of only two sentences when the text is complete. The first sentence of the section, which is the last but one sentence of the text as a whole is the one that was taken out to provide the item. A correct response requires reading, apart from the (sentence in the) given section, also some other sections of the text and understanding some of the main ideas in those sections as well. There is some amount of lexical overlap between the particular section of the text in which the item is located and the correct answer ('*stay away*' is used in the section, '*stay with*' in the correct answer).

TASK 3 (*Gorillas in …*)

Text type: A descriptive-narrative text from a newspaper

Item type: Matching headings to paragraphs of a text

Reading (sub)skills tested: ability to understand the gist of a passage

Item 1 EV: *v7, v14, v19*

It is possible to answer the item correctly by reading only the two-sentence long paragraph of the text in focus of the item. A correct answer requires understanding the main idea (information about the place where the gorillas live), and recognizing the relationship between the meaning of the paragraph and the correct answer (the suitable paragraph heading) (Option D, '*The location*'). The paragraph has no lexical overlap with the correct option, whilst it has with an incorrect one. (The last word of the paragraph '*population*' is used in Option A, '*How the gorilla population is organised*', which is, however, not the answer to this item.) Although the paragraph includes a relatively high number of lower-frequency words, responding to the item is relatively easy, because the crucial information is expressed with vocabulary items likely to be familiar even to lower level students (e.g., '*forest in the south-west*', '*National Park*', '*is home to … 300 mountain gorillas*'.)

Item 2 EV: *v1, v3, v7, v12, v14, v19*

A correct answer requires reading only the three-sentence long paragraph in focus of the item. It involves understanding the main idea in the paragraph, and recognizing the relationship between the paragraph and the suitable paragraph heading (Option A, '*How the gorilla population is organised*'). Two of the three sentences of the paragraph are syntactically complex sentences, and both use the passive voice. The passive structure, also used in the correct answer, is likely to be difficult to understand for lower level students. The paragraph contains some difficult or less frequently used words and

phrases (e.g., '*split into*', '*habituated*', '*consists of*', '*troop*', '*human presence*'). However, it is possible to answer the item by understanding easier phrases in the paragraph, like '*23 groups*', or '*13 animals*'. While the correct option has no lexical overlap with the paragraph, two incorrect options do. Specifically, the word '*group*', used twice in the paragraph, appears in Options C ('*The leader of the group*') and H ('*What the leader of the group did*'), neither of which is the correct answer to this item.

Item 3   EV: *v1, v3, v7, v10, v14, v17*

A correct answer can be given by reading only the three-sentence long target paragraph and understanding the main idea in it. Two of the three sentences of the paragraph are syntactically complex sentences, one of which uses the passive voice ('*Six females and six young are led by …*'). The paragraph has lexical overlap with the correct option (Option C, '*The leader of the group*'), and also with an incorrect one (Option H, '*What the leader of the group did*'). The key word '*leader*' in the correct option might be unfamiliar to lower-level students.

Item 4   EV: *v4, v7, v10, v15, v20*

A correct answer requires reading not only the two-sentence long 'target' paragraph, but also the paragraph preceding it. It requires understanding the main ideas in, and recognizing meaning relationships (relatively simple anaphoric reference relations) between the two paragraphs. Sentences of the target paragraph (Paragraph 4) contain quite a few lower-frequency words and phrases ('*munched*', '*contentedly*', '*vegetation*', '*displaying his 8ft reach*', '*sustenance*'), some of which might be crucial for understanding the gist of the paragraph. A key word ('*leader*') in the correct answer (Option H, '*What the leader of the group did*') might be unfamiliar to lower-level students. An incorrect option, talking about the gorillas' '*reaction*', has a degree of

semantic overlap with the correct option. Its elimination requires comparing the two options, and checking the meaning of each against the meaning of the paragraph.

Item 5   EV: *v3, v4, v5, v7, v10, v14, v16*

It is possible to answer the item correctly by reading only the four-sentence long paragraph in focus of the item. However, the topic of the paragraph is abstract, which is likely to make the main idea more difficult to understand. Most sentences are syntactically simple sentences. There is only one complex sentence, which is, however, relatively long (28 words) and, in addition, uses a passive structure. The paragraph contains quite a number of more difficult words and phrases like '*primeval feelings*', '*crouching*', '*instincts*', '*tremendous*', '*edge of extinction*', '*privilege*', some of which might be crucial for understanding the main idea. The meaning of the correct answer is likely to be difficult for lower level students to understand, because of the vocabulary used in it (Option E, '*Appreciation of a unique experience*'). There is, however, a degree of lexical overlap between the paragraph and the correct answer. (The word '*impressive*' is used in the first sentence of the paragraph.)

Item 6   EV: *v3, v4, v5, v7, v12, v14, v16*

It is possible to answer the item correctly by reading only the six-sentence long target paragraph, which is the longest paragraph of the text. Its topic is slightly more abstract than the topic covered in most other paragraphs. While most sentences are syntactically simple sentences, three of them involve the passive voice. The paragraph uses many lower-frequency words and phrases (e.g., '*to such an end*', '*currently engaged*', '*money .. ploughed back into*', '*revenue sharing*', '*hangs by a thread*'), some of which might be essential for understanding the main idea of the paragraph. Apart from the paragraph itself, the correct answer also uses the passive voice ('*What is done to protect the gorillas*'), which might cause difficulty to lower level students. However, there is 'exact

97

word' lexical overlap between the given section of the text and the correct answer, with the word '*protect*' used in both, which makes the selection of the correct answer possible even for lower-level students.

TASK 4 (*Being wet …*)

Text type: A narrative of personal experience taken from a teenage magazine

Item type: Matching sentences to gaps in text

Reading (sub)skills tested: ability to understand main ideas

Item 1   EV: *v1, v7, v10, v15, v17, v21*

The item requires processing not only the particular two-sentence long section of the text in which the item is located but also the preceding section. It requires understanding the main ideas as well as the relationship between the two sections. Although both sentences of the section gapped to provide the item are syntactically complex sentences, the crucial information is easy to understand ('*We weren't far from the station when ..*'). The correct answer to the item is the shortest, and grammatically easiest, sentence among the 9 options provided ('*Eventually we wandered back to catch the 2 pm train home.*'). However, key vocabulary for understanding its meaning ('*eventually*', and '*wandered back*') might be unfamiliar to lower level students. The section has lexical overlap with not only the correct option, but also with four incorrect options. The elimination of the incorrect options requires comparing their meaning and making an inference which is a more suitable answer to the item. In addition, the distractor is a very plausible option if one reads only the target section of the text. Its elimination requires reading and understanding the information in, apart from the gapped section and the section that precedes it, a further section of the text. As the item is the first item

on the task, it may also require first skimming through the whole text, along with the options, to get an overall idea of the main narrative events in the text.

Item 2   EV: *v1, v7, v9, v11, v12, v14, v17, v21*

It is possible to answer the item correctly by reading only the two sentences that precede the gap for the item. Although both sentences are complex, the main idea is relatively easy to understand ('*the sky went black*', '*there was a huge* [clap] *of thunder*', '*the rain came down so hard..*'). However, the correct answer to the item is somewhat long (21 words) and uses, on the one hand, a lower-frequency word that carries crucial information ('*drenched*') and, on the other, the passive structure, which is likely to be unfamiliar to lower level students ('*we were caught in it*', '*we were drenched*'). The item has lexical overlap with the correct option, as well as with four incorrect options. Three of these incorrect options can be eliminated by comparing their meaning, and checking the meaning of each against the main idea of the section. The elimination of the fourth (Option E, '*One thing is for sure, though, we're all taking umbrellas next time we go shopping.*') requires reading various other sections of the text and making inferences based on information in those sections.

Item 3   EV: *v1, v2, v7, v10, v12, v15, v18, v21*

The item requires reading two consecutive sections of the text. Most sentences involved are relatively long and complex sentences. The main idea in the section in focus of the item is relatively easy to understand. However, the answer to the item (Option D, '*My friends and I were too shocked to argue, so we just let the train leave the station.*') uses a passive structure and verbal phrases that may be difficult for lower-level students ('*were shocked to argue*', '*let the train leave*'). The section has lexical overlap with the correct option and three incorrect ones. The elimination of the three incorrect options is

relatively easy, because the overlap with the correct option is greater. Among the options, there is one very plausible option, specifically, Option A, the answer to the next item on the task (Item 4), whose elimination is much more difficult. It requires reading a further section of the text and making inferences based on information in the sections read and the options involved.

Item 4   EV: *v1, v3, v4, v7, v11, v12, v15, v21*

The item requires reading three sections of the text: the one in which the item is located, and those immediately preceding and following it. Sentences of the target section are complex sentences and use the passive voice. Some phrases in the section are likely to make the main idea more difficult to understand than in other sections of the text (e.g., 'watch [..] *other passengers pull away*', '*they'd all avoided*', '*were being picked on*'). The correct answer (Option A) is the longest of the 9 options (32 words), and also it appears to be the most difficult in terms of grammatical structures ('*We'd have been happy to stand if they were worried we'd wreck the seats, but now we had to ..*') . An incorrect option (Option D) is a very plausible option. Its elimination requires making inferences based on information in at least three different sections of the text.

Item 5   EV: *v1, v2, v7, v11, v15, v18*

The item requires reading and understanding the main ideas in the one-sentence long section that includes the gap and the first sentence of the section that immediately follows the gap. It may also be necessary to understand some information in the section that precedes the one in focus of the item. The sentences involved are long and complex sentences, the crucial information, however, is relatively easy to understand ('*We sat around freezing cold*', '*the next train came*', '*no problem getting on*'). The answer to the item (the suitable sentence for the gap) is relatively long (26 words). The section has

lexical overlap with the correct answer as well as with some incorrect options. However, the overlap with the correct option is greater ('*freezing cold*' is used in the text, '*shaking with cold*', '*got [..] into the bath to warm up*' in the correct answer).

Item 6 EV: *v1, v2, v7, v11, v14, v16*

It is possible to answer the item correctly by reading only the two-sentence long section of the text that includes the gap for the item. Although both sentences are long (21 and 30 words) and syntactically complex sentences, the crucial information for a correct answer is easy to understand ('*When I told my mum what [..] happened*', '*she [..] rang up South West Trains*'). The correct answer is also long (29 words) and contains some difficult grammatical structures. However, an understanding of those structures is not necessary for students to be able to select the correct answer, because there is exact word lexical overlap between the given section of text and the correct answer (Option B, '*All my mates' mums wrote to the train company, asking if ..*').

Item 7 EV: *v7, v15, v21*

The item is located in the last section of the text, which consists of two sentences when the text is complete. Of the two sentences of the section, the second sentence was taken out to provide the item. Given that the item involves the very last, concluding sentence of the text ('*One thing is for sure, though, we're all taking umbrellas next time we go shopping.*'), selection of the correct answer requires reading and understanding the main ideas in most sections of the text. If one reads and understands the meaning of only the section in which the item is located, many of the incorrect options might seem to be possible answers.

TASK 5 (*Caught out in the rain*)

Text type: A mainly narrative text taken from a newspaper

Item type: Matching headings to paragraphs of a text

Reading (sub)skills tested: ability to understand the gist of a passage

Item 1   EV: *v1, v2, v4, v7, v9, v15, v19, v21*

The item requires reading, apart from the paragraph in focus of the item, also the preceding paragraph, which is the introductory paragraph of the text and provides the example. A correct answer requires understanding the main ideas as well as the relationships across sentences of the two paragraphs. Most sentences involved are long and complex sentences using many lower-frequency vocabulary items, some of which might be essential for understanding the main ideas in the paragraphs in question (e.g., '*vast sewer reconstruction scheme*', '*retrace my steps*', '*make a [..] detour*', '*salvation seemed at hand*', '*loomed up on*', '*glancing through*'). The correct answer (the suitable paragraph heading) consists of two words (Option G, '*Possible short-cut*'), one of which is a lower-frequency word ('*short-cut*') that might cause difficulty even for some higher level students. It appears to contribute to the difficulty of identifying the correct answer that if one only reads the paragraph in focus of the item, many of the 8 options may seem to be a possible answer to the item. In particular, the distractor (Option D, '*The best way to find shelter from the rain*') is very plausible if one reads the introductory paragraph of the text superficially or does not understand certain details in that paragraph. To be able to eliminate all superficially plausible but incorrect options, students may need to read also some of the other paragraphs of the text. In addition, the paragraph has lexical overlap with two incorrect options, but not with the correct option. The second word of the paragraph, '*suddenly*', appears in Option F ('*A sudden obstacle*'), while '*office building*' mentioned in the paragraph appears in Option C

('*Two approaches to public use of office buildings*'), neither of which is the answer to the item.

Item 2   EV: *v1, v2, v4, v7, v9, v15, v21*

The item requires processing, apart from the paragraph in focus of the item, two further paragraphs, specifically, those immediately preceding and following the passage. If one only reads the paragraph in focus of the item, some of the incorrect options might seem to be possible answers to the item. Sentences of the target paragraph are long and syntactically complex sentences. One of them is a 49-word long sentence with six sub-clauses. The paragraph contains some phrases, idiomatic expressions that are likely to make processing the sentences involved more difficult (e.g., '*steely expression*', '*be on one's mind*', '*be up to*', '*a rat run*'). Of the two content words of the correct answer (Option F, '*A sudden obstacle*') one is a lower-frequency word ('*obstacle*'), whose knowledge is likely to be crucial for understanding the meaning of the given option and identifying it as the correct answer.

Item 3   EV: *v1, v7, v10, v15, v16, v21*

It is possible to answer the item correctly by reading only the paragraph in focus of the item. To be certain of the correct answer and eliminate incorrect options that are plausible if we do not read any other paragraphs of the text, one needs to read at least one more, the preceding, paragraph. Although most sentences involved are syntactically complex, they are relative short and use simple grammatical structures. Crucial information is easy to understand because most of the vocabulary in the paragraph consists of relatively easy, high frequency words ('*I have an appointment with Mr Henderson, I lied.*', '*I think he's on the first floor.*', '*Just a minute.*'). The correct answer to the item is fairly easy to understand (Option A, '*A trick – will it fail?*'), although both '*trick*' and '*fail*' might be unfamiliar to lower level students. The

paragraph has easily recognizable lexical overlap with the correct answer ('*lied*' used in the text, '*trick*' in the correct answer).

Item 4   EV: *v3, v7, v9, v15, v17, v21*

The item requires processing at least two consecutive paragraphs of the text, the one in focus of the item and the paragraph that precedes it. It involves understanding the main ideas, as well as the relations across the sentences involved. Most sentences of the target paragraph are short, simple sentences, using for the most part easy, high-frequency words and phrases ('*phone*', '*desk*', '*answer*', '*be off*', '*stairs*', '*street*', '*rain*'). One of the key words, however, is included in a passive structure ('*I was saved*'). While the main idea in the paragraph is relatively easy to understand, the answer to the item (Option B, '*An unexpected narrow escape*') uses lower-frequency vocabulary, which makes understanding its meaning more difficult than is the case with the main idea in the text. If one reads only the paragraph in focus of the item, more than one of the options may seem to be a possible answer to the item. Their elimination requires reading some of the other paragraphs of the text as well. Furthermore, the item has lexical overlap with the correct answer and two incorrect options (Options D and E). They are, however, easy to eliminate on the basis of their easily understandable content.

Item 5   EV: *v1, v2, v3, v4, v5, v7, v10, v14, v16*

It is possible to answer the item correctly by reading and understanding the main idea of the paragraph in the focus of the item. The paragraph is the longest of the seven paragraphs of the text, although it consists of only four sentences. Three of the four sentences are complex sentences, using both fairly complex grammatical structures and occasionally rather long phrases involving low frequency words as well ('*reflect credit on*', 'a*dminister ordinary commercial office buildings as though*', '*outweighed by*',

'*accrues*', '*corporate image*'). The content covered in the paragraph is mainly abstract and, as a result, it is more difficult to recognize the semantic relationships across the sentences involved and identify the main idea. The answer to the item uses vocabulary that is likely to cause difficulty to lower level students (Option C, '*Two approaches to public use of office buildings*'). There is, however, "exact word" lexical overlap between the paragraph and the correct answer (the words '*office buildings*' and '*public*' occurring in the paragraph are used in the correct answer).

Item 6   EV: *v4, v5, v7, v14, v17*

It is possible to answer the item correctly by processing only the short, two-sentence long paragraph in focus of the item. In fact, the crucial information for a correct answer is in the first sentence of the paragraph, which is a syntactically complex sentence. Although some of the vocabulary might cause difficulty even to some higher level students (e.g., '*transferred*', '*rural*' or '*attitude*'), the phrase '*Get off my land*' in the section is relatively easy to understand. The topic of the paragraph, similarly to the previous item, is abstract. The correct answer involves a saying, which is, however, easily understandable even for lower level students (Option E, '*An Englishman's home is his castle*'). The paragraph has a degree of lexical overlap with the correct answer (the paragraph begins with the words '*Here in Britain*'), and an incorrect option (the word '*approach*' used in the paragraph appears in Option C, which is not the answer to this item).

TASK 6 (*Animals under threat*)

Text type: A mainly expository text from a magazine

Item type: Matching clauses to gaps in text

Reading (sub)skills tested: ability to understand text structure

Item 1   EV: *v2, v5, v8, v10, v11, v15, v20, v22*

It is possible to answer the item correctly by processing only the gapped sentence. It involves recognizing syntactic and semantic relationships between the gapped sentence and the correct response (the clause that fits the gap). The four-word long first part of the gapped sentence ('*Unless we act now*,') uses a word (the conjunction 'unless') that is likely to be unfamiliar to lower level students, while its understanding is crucial for identifying the correct answer. The correct answer (the main clause of the gapped sentence) is the longest of the ten options (Option E, 21 words). It uses easy grammatical structures, but includes some difficult vocabulary items, whose knowledge, however, is not crucial for a correct response. Apart from the correct option, an incorrect option (Option C) is very plausible on the basis of its content. Its elimination requires comparing the two options in terms of syntactic features and checking each option against the gapped sentence. As the superficially plausible but incorrect option (Option C) is the correct answer to the next item in the task (Item 2), it may be necessary for students to read, apart from the gapped sentence, also the first or first two sentences of the section that follows the item, where Item 2 is located, answer Item 2 first, thereby eliminating Option C as a possible answer to this item. It is possible to eliminate most incorrect options by recognizing their syntactic inappropriateness.

Item 2   EV: *v2, v5, v8, v9, v15, v20, v22*

It is possible to answer the item correctly by reading only the gapped sentence. It requires recognizing syntactic and semantic relationships between the first part of the gapped sentence ('*the fact is that*') and the clause providing the correct answer (Option C). However, apart from the correct answer, two incorrect options (Option E, the correct answer to the previous item, Item 1, and Option J) are very plausible on the basis of

their content. Their elimination requires comparing the options involved and making an inference which is more suitable. It may also be helpful for students to read, in addition to the gapped sentence, some other sections of the text, including, above all, the four-sentence long introductory paragraph of the text that precedes the item. The vocabulary used in the correct answer involves two lower-frequency words ('*species*' and '*extinct*') whose meaning, however, may become clear from the text. Most incorrect options can be eliminated on the basis of syntactic features.

Item 3   EV: *v2, v4, v5, v8, v9, v14, v19*

The item requires processing only the gapped sentence. It involves recognizing syntactic and semantic relationships that the correct option (Option I) has with the gapped sentence. The gapped sentence is a 24-word long complex sentence, whose middle part (an adverbial clause, beginning with "*just so that*") is taken out of the text to provide the item. There is lower-frequency vocabulary in the crucial information in both the gapped sentence ('*pacing up and down*') and the correct answer ('*just so that*' and '*gape at*'). An important source of difficulty for the item is that the correct option needs to be checked against both the beginning and the end of the gapped sentence, in terms of both syntactic features and content. The item has lexical overlap with an incorrect option, but not with the correct option. The elimination of most incorrect options requires a detailed understanding of their content.

Item 4   EV: *v2, v5, v8, v14, v22*

The item can be answered correctly by processing only the gapped sentence. The gapped sentence is a complex sentence, involving a simple defining relative clause, with easy grammatical structures. The correct answer is the middle part of the gapped sentence. All eight incorrect options can be eliminated relatively easily on the basis of

both syntactic features and their inappropriate content, even though some of them may require more careful attention than the others, because of either syntactic or content features.

Item 5   EV: *v2, v5, v8, v9, v14, v22*

It is possible to answer the item correctly by processing only the gapped sentence. The sentence is a syntactically simple sentence, which, however, involves a relatively long infinitive phrase, used as a postmodifier to the comparative adjective "better" ('*Wouldn't it be better to pour the time and money into preserving these animals in their natural habitats?*'). The infinitive phrase was taken out of the text to provide the item. The vocabulary in the correct answer includes lower-frequency words ('preserve', 'habitats'). Most incorrect options can be eliminated relatively easily on the basis of syntactic features.

Item 6   EV: *v4, v5, v8, v9, v15, v21*

It is possible to answer the item correctly by processing only the gapped sentence. However, to be certain of the correct answer, students may need to read one or two sentences before and/or after the item. The gapped sentence itself is a long and complex sentence (28 words when complete) and uses lower-frequency vocabulary, including some topic-specific words and phrases ('*clutch of eggs*', '*hatchlings*'), whose understanding may be crucial for a correct answer. Both the syntactic complexity of the sentence and the topic-specific vocabulary involved are likely to make understanding the content of the sentence relatively difficult even for many higher level students. The phrase '*genetic make-up*' in the correct answer may cause difficulty for lower level students. Of the eight incorrect options, only four can be eliminated easily on the basis of syntactic features (Options C, D, E, and J). One incorrect option (Option F) is a syntactically possible answer and is plausible in terms of its content as well. Its

elimination requires comparing its content against the content of both the correct option and some of the sentences preceding and following the item, and making an inference.

Item 7   EV: *v5, v8, v12, v15, v21*

The item can be answered correctly by processing only the gapped sentence. However, to be certain of the correct answer, students may need to read sentences in various other sections of the text. Although the gapped sentence itself is a complex sentence, the grammatical structures involved in its main clause in the input text are not particularly difficult ('*they will be able to reach the higher leaves, [..] and so survive*'). The correct answer (Option D, '*which haven't yet been eaten*') uses the Present Perfect tense in the Passive, which is likely to be unfamiliar to lower level students. The vocabulary used in both the gapped sentence and the correct answer is easy, although the word '*leaves*', which is crucial for understanding the meaning of the gapped sentence, might cause difficulty to lower level students. Of the eight incorrect options, only three can be eliminated very easily on the basis of syntactic features (Options C, E, and J). Two of the syntactically possible options are plausible in terms of their content as well (Options B and F). Their elimination requires comparing their content against the content of the correct answer, on the one hand, and the content of both the gapped sentence and some of the sentences before and after the gapped sentence, on the other, and making an inference.

Item 8   EV: *v1, v2, v5, v8, v9, v14, v17, v22*

The item requires processing only the gapped sentence. The sentence is a relatively long (22 words) and complex sentence. The vocabulary in the first part of the sentence (in the text) consists of easy, high frequency words ('*In fact, of all the animals which have lived on earth*'), while the correct answer (Option J) involves two lower-frequency words ('*evolved*' and '*extinct*'), whose meaning, however, may become clear from the

text. It is easy to eliminate most (seven of the eight) incorrect options on the basis of syntactic features. The only syntactically possible option (Option E) can be eliminated by comparing its content against the content of the correct answer. However, the item has lexical overlap with the correct answer and an incorrect option (Option C). (The word '*extinction*', used in the sentence immediately preceding the gapped sentence, appears in an adjectival form '*extinct*' in both Option J, the correct answer, and Option C, an incorrect option.)

4.4 Summary of the results

The tables below (Tables 4.6 and 4.7) provide a summary of the results related to characteristics of the tasks and items identified through content analysis. Table 4.6 shows the frequency of occurrence of each variable in the set of 42 items analysed, whilst Table 4.7 shows the distribution of variables across items and tasks, on the one hand, and the most frequently occurring variables by task, on the other.

Table 4.6 Frequency of occurrence of each variable

Variable:	**V7**	**v14**	**v16**	**v15**	**v1**	**v10**	**v2**	**v5**	**v21**	**v4**	**v6**
# of items:	24	17	16	15	13	13	12	12	11	10	10

Variable:	**v13**	**v9**	**v8**	**v17**	**v3**	**v11**	**v12**	**v22**	**v18**	**v19**	**v20**
# of items:	10	9	8	8	7	6	6	5	4	4	3

Table 4.7 Distribution of variables across tasks and items

Task No	Item No	Variables in each item	Most frequent variables by task
T1	**Item 1**	v6, v10, v13, v17	v6 – 10 items
	Item 2	v6, v13, v16	v13 – 10 items
	Item 3	v6, v13, v16	--------------------

	Item 4	v6, v13, v16	v16 – 7 items
	Item 5	v6, v13, v16	
	Item 6	v6, v13, v16	
	Item 7	v6, v13, v16	
	Item 8	v6, v10, v13, v18	
	Item 9	v6, v10, v13, v16	
	Item 1	v6, v10, v13, 17	
T2	Item 11	v7, v14, v16	v7 – 5 items
	Item 12	v7, v14, v16	v14 – 4 items
	Item 13	v7, v10, v11, v14, v16	------------------
	Item 14	v7, v14, v18	v16 – 4 items
	Item 15	v7, v15, v16	
T3	Item 16	v7, v14, v19	v7 – 6 items
	Item 17	v1, v3, v7, v12, v14, v19	v14 – 5 items
	Item 18	v1, v3, v7, v10, v14, v17	------------------
	Item 19	v4, v7, v10, v15, v20	v3 – 4 items
	Item 20	v3, v4, v5, v7, v10, v14, v16	
	Item 21	v3, v4, v5, v7, v12, v14, v16	
T4	Item 22	v1, v7, v10, v15, v17, v21	v7 – 7 items
	Item 23	v1, v7, v9, v11, v12, v14, v17, v21	v15 – 5 items
	Item 24	v1, v2, v7, v10, v12, v15, v18, v21	------------------
	Item 25	v1, v3, v4, v7, v11, v12, v15, v21	v1 – 6 items
	Item 26	v1, v2, v7, v11, v15, v18	v21 – 5 items
	Item 27	v1, v2, v7, v11, v14, v16	v11 – 4 items
	Item 28	v7, v15 v21	
T5	Item 29	v1, v2, v4, v7, v9, v15, v19, v21	v7 – 6 items
	Item 30	v1, v2, v4, v7, v9, v15, v21	v15 – 4 items
	Item 31	v1, v7, v10, v15, v16, v21	------------------
	Item 32	v3, v7, v9, v15, v17, v21	v1 – 4 items
	Item 33	v1, v2, v3, v4, v5, v7, v10, v14, v16	v4 – 4 items
	Item 34	v4, v5, v7, v14, v17	v21 – 4 items
T6	Item 35	v2, v5, v8, v10, v11, v15, v20, v22	v8 – 8 items
	Item 36	v2, v5, v8, v9, v15, v20, v22	v14 – 4 items
	Item 37	v2, v4, v5, v8, v9, v14, v19	v15 – 4 items
	Item 38	v2, v5, v8, v14, v22	--------------------
	Item 39	v2, v5, v8, v9, v14, v22	v5 – 8 items
	Item 40	v4, v5, v8, v9, v15, v21	v2 – 6 items
	Item 41	v5, v8, v12, v15, v21	v9 – 5 items
	Item 42	v1, v2, v5, v8, v9, v14, v17, v22	v22- 5 items

As figures for the three 'item type' related variables (v6, v7 and v8) in Table 4.6 indicate, most items examined were main idea items (v7 – 24 items), ten were classified as requiring the ability to locate specific information (v6), and eight measuring the ability to understand information and recognize structural relations within the sentence (v8). Among the most frequently occurring variables are v14 (17 cases), v15 (15 cases),

and v16 (16 cases). The first two of these are related to the amount of processing supposedly needed for a correct answer, while the third one is a lexical overlap variable. As can be seen from Table 4.7, v14 characterizes most items in Tasks 2 and 3, while most items in Tasks 4 and 5 involve v15. This means that, according to the analysis, most items in Tasks 2 and 3 can be answered correctly by reading one specific section of the text, whereas for a correct answer to most items in Tasks 4 and 5 students may need to read and understand information across two or more sections of the text. With respect to the most frequent 'lexical overlap' variable (v16: 'The item has a lexical overlap with the correct option but not with the other options'), Table 4.7 shows that the vast majority of its occurrence (11 out of 16 cases total) involve items in Tasks 1 and 2 (seven out of ten items in the case of Task 1, and four out of five items in the case of Task 2), which suggests that v16 is an important item characteristic in the case of these two tasks.

The most frequently occurring text-related variable is v1 ('Most sentences of the target section of the text are syntactically complex sentences'), involving 13 cases total. In most cases (10 out of 13), it occurs with items in Tasks 4 and 5, supposedly increasing the difficulty of processing the items involved. Two other relatively frequent text-related variables are v2 ('Sentences of the target section tend to be long') and v5 ('The content of the target section of the text is abstract'), involving 12 cases each. Notable is the fact that both are among the most frequently occurring variables in the description of items in Task 6, but not in the rest of the tasks. The total number of cases identified for vocabulary-related variables (v4, v9, and v10) is 32. Of these, v4 ('There is lower-frequency vocabulary in the crucial information in the text') characterizes most (4 out of 6) items in Task 5, v9 ('Key vocabulary for understanding the question or the correct

answer includes lower-frequency words') was identified to play an important part in answering most (5 out of 8) items in Task 6, whilst the occurrence of v10 ('Key vocabulary for understanding the question or the correct answer includes words that might be unfamiliar to lower-level students') is scattered across the tasks and items analysed. Finally, in the category of variables related to the elimination of plausible but incorrect options, v21 ('The elimination of incorrect options requires an inference based on information in different sections of the text') was observed to characterize more items than v20 ('The elimination of incorrect options requires comparing the meaning of options and making an inference based on information in the given section of the text') (11 vs 3 items). The vast majority of cases for v21 (nine out of eleven) involve items in Tasks 4 and 5.

4.5 Concluding remarks

In this chapter we have described the tasks and items under investigation employing the methodology of content analysis. We have identified 22 variables under eight broader categories of item characteristics likely to affect the difficulty of answering these reading items. One of the most important findings of the study is related to the complexity of processing these items. The analysis has shown that there are a number of variables underlying performance on any of the items examined. It has also shown that most item characteristic variables identified are not evenly distributed across the tasks and items. Certain variables have been found to characterize most items in one particular task but none of the items in another task, other variables characterize most items in two other tasks, and the occurrence of yet other variables appears to be shared by certain items in most of the six tasks examined. All this suggests considerable

variation in the construct measured by the six reading tasks analysed. The next chapter will examine characteristics of the same set of items using the methodology of verbal protocol analysis, to which we will turn now.

Chapter 5 Study Two: Verbal Protocol Analysis

5.1 Introduction

In language testing, the principal questions centre on issues of validity and more recently verbal protocols have been increasingly used as part of construct validation. Protocol analysis allows us to better understand what test-takers actually do when they take tests, hence, what it is that a given test actually measures, therefore, it may be a useful tool for providing evidence for construct validity by supplementing data obtained from more conventional quantitative techniques. Furthermore, it may be usefully applied in the process of test development to select, evaluate and modify test materials, including the development of specifications, commissioning of test tasks, selection and editing of tasks, trialling, pretesting and item analysis. It may play a part in helping to specify the characteristics of the task, items, or ability in question, which might be difficult to reveal with less direct methods (Green 1998). In general, data from verbal protocols can supplement information on the task or items provided by quantitative methods or formal task analysis.

For the purpose of this research, it was decided to use VPA, because it was considered important to explore the difficulty of the tasks and items under investigation also from a perspective rather different from that offered by either the content analysis approach used in Chapter 4 or the empirical indicators of item difficulty looked at in Chapter 3, specifically, the perspective of the test-taker. In particular, this study (Study Two) was intended to examine how L2 learners go about processing the matching tasks focused on in this research, how they actually respond to these reading items as opposed to what

they may be expected to be doing when completing the items. The main research question this study sought to answer was formulated as follows:

RQ 2: *What skills, knowledge and processes are used by test-takers to complete the reading items under investigation?*

5.2 Methodology

Participants

Altogether six secondary school students, aged 16-19, were involved in the study. At the time of the data collection, they were all studying in the same secondary school of art in Pécs, Hungary. One was from Year 10 (LMS), two from Year 11 (LS and MS), one from Year 12 (HS2) and two from Year 13 (MHS and HS1), and all six were females. Five of them started learning English as a foreign language in primary school, while one student (MHS) studied German in primary school and started learning English in the second form of secondary school, after passing an international intermediate level language examination in German. The Year 10-12 students all had four normal English lessons a week, while the two Year 13 students, after having taken their final examination at the end of Year 12, had the opportunity to attend only two lessons a week to prepare for a language examination. Both high-level students (HS1 and HS2) had passed their intermediate language examination in English a few months before data collection started. Three of the six participants were selected from language groups taught by the researcher, and three from a colleague's groups in the same school. The students' language level, as judged by their teachers, ranged from low to high (hence their codes: LS=Low-level Student; LMS=Low/Middle-level Student;

MS=Middle-level Student; MHS=Middle/High-level Student; HS1 and HS2=High-level
Student 1 and 2).

Selection criteria:

In the selection process, the language level of the students was considered important
only to the extent that the tasks they were to complete were of differing levels of
difficulty. That is, it was important to involve informants of both higher and lower
levels of language knowledge, but it was not felt necessary to employ a special
measurement instrument to obtain an exact measure of the difference between
participants' language levels, since the main focus was not the difference between test-
taking processes and strategies used specifically by successful vs unsuccessful learners,
but rather the skills used and the sorts of difficulties encountered by students, whether
of higher or lower levels, when answering the matching items investigated in the study.
For the purpose of this research, the most important criteria for selecting participants
were their ability to express their thoughts, to 'think aloud' while completing a task, on
the one hand, and their willingness to participate in the research, on the other. They
were not intended to constitute a representative sample, nor was the selection intended
to be random, from which to make generalizations. The main selection principle was
that, as Alderson (1990:468) points out, 'in qualitative research of this kind, it is more
important to identify good informants than to find representative informants'.

Every effort was made to meet the above selection criteria. It was relatively
straightforward to identify suitable informants from among students in the language
groups taught by the researcher. To ensure that the above criteria were met in the case
of the remaining informants as well, the colleague, who kindly agreed to help, was
informed about the purpose of the study and crucial aspects of the data collection

procedure, and had also been given a detailed explanation of the selection criteria, before attempting to identify suitable subjects. The students' ability to verbalise their thoughts was assessed by their teachers, relying on their experience and observation of the students' ability to talk about the difficulties they encountered while completing tasks, in general, and reading tasks, in particular, in the classroom context.

Materials

Materials to be prepared and equipment for use in the data collection session included the following:

- Task sheets (for subjects to work on during the data collection session)
- A copy of each task for the researcher (to make it easier to follow how the subject is processing the task)
- A questionnaire (to be completed by subjects after the session)
- A tape recorder for recording the session

For details about the tasks involved and the questionnaire, see the section below.

Data collection design

At the initial stages of designing data collection, it was planned to include in the VPA study, if possible, all eight tasks that were originally examined for their content characteristics in the early phases of our investigation. The main guiding principles for designing data collection were the following. Each subject was to complete two or three tasks, depending on the difficulty of the tasks they were to complete, their language level, their ability to produce a verbal report, and the envisaged effect of the interaction of these with the length of the session. On the other hand, each task was to be completed by two students, a lower level and a higher level student. The plan also included the

possibility that, in the case of some tasks, where insights from the content analysis study (Chapter 4) suggested, a third informant might be involved. However, within the time frame available for carrying out this study, it was possible to collect only one protocol each in the case of two of the eight tasks, therefore, it was decided to drop those two tasks from the rest of the procedure. Thus, the resulting data set consisted of 15 protocols generated on six tasks, as shown in the table below.

Table 5.1 Data collection design

Task No	Task ID	Students completing each task		
T1	**Julie wants** (MST)	**LS**	**LMS**	
T2	**Giant pandas** (MSG)	**MS**	**HS1**	**HS2**
T3	**Gorillas in Uganda** (MH)	**LS**	**LMS**	
T4	**Being wet** (MSG)	**MS**	**HS1**	**LS**
T5	**Caught out in the rain** (MH)	**LMS**	**HS1**	**HS2**
T6	**Animals under threat** (MCG)	**MHS**	**HS2**	

Codes: LS=Low-level Student; LMS=Low/Middle-level Student; MS=Middle-level Student; MHS=Middle/High-level Student; HS1=High-level Student 1; HS2=High-level Student 2; MST=Matching Statements to text; MSG=Matching Sentences to Gaps in text; MH=Matching Headings to text; MCG=Matching Clauses to Gaps in text;

Of the 15 verbal reports gathered, three had to be discarded for one or the other of the following two main reasons. First, the length of the data collection session during which the given protocol was collected, despite careful planning, exceeded well beyond the ideal time limit of one hour, and the task in question was completed by the student as her third task in the second half of the session, when exhaustion was likely interfere, to a great extent, with the student's ability to concentrate on the task (Table 5.1: T4 by LS). Second, within the time available for the study, it was not possible to carry out a careful analysis of the protocol (Table 5.1: T2 by HS2, and T5 by HS2). This resulted in

a total of 12 protocols for the final analysis. (For their identification, see the shaded cells in Table 5.1 above.)

In order to complement the verbal report data with a different type of information on the same topic, a follow-up questionnaire was devised, asking subjects to assess the difficulty of each task and each item they responded to, on a 1-6 scale of difficulty. (A sample questionnaire is given in Appendix B.)

Procedure

After identifying subjects, individual appointments were made with each a few days before the actual data collection session. This meeting was to inform participants about the more general aim of the research they were asked to participate in, introduce them to the concurrent verbal report procedure, explaining what is meant by thinking aloud while carrying out a reading task and what they were expected to do during the think-aloud session. They were also told that the session was likely to take approximately an hour, and that they could withdraw from the study if, in the light of detailed information about what was required of them, they wished to do so. However, all selected subjects were very motivated and interested, some even asked questions related to further details of not only the think aloud procedure itself, but the research in general. At the end of the meeting, the exact time and date of the individual think-aloud sessions was set with each subject.

The think-aloud session consisted of three main phases, as described below.

Phase 1: First, students were reminded of the way they were required to carry out the tasks. They all remembered what they had been told about the procedure at the meeting prior to the session. They were then provided an opportunity to practise. For this

purpose, two short practice tasks taken from Green (1998) were used. One asks subjects to think aloud as they multiply two numbers in their head, the other to think aloud as they add up all the windows in their house. At the end of each practice task, students were given feedback to let them know if they had produced their reports as required. During these tasks, they were interrupted if their report did not reflect their thoughts as they were processing the task, and were given prompts to help them produce the report appropriately. In general, the two tasks proved to have provided sufficient training for them, as is evidenced in the reports generated on the reading tasks in Phase 2 of the think-aloud session.

Phase 2: In the main part of the session, students were asked to complete their reading tasks in the same way as they had completed the practice tasks, i.e., thinking aloud. With the exception of one informant (MHS), all students were asked to carry out three reading tasks within a roughly 60-minute time frame. The reason for assigning only two tasks for the Middle/High student was that, due to the difficulty level of the tasks, as well as the student's language level, completion of the two tasks was anticipated to fill the planned time frame of the session. This, in fact, proved to be the case during administration, as she spent 57 minutes carrying out the two tasks. The task sheets were handed over to students one by one, that is, when they finished working on one task, the completed task sheet was collected and the next task handed over. In this phase of the session, students were, for the most part, only interrupted when they fell silent or talked in a very quiet, hardly intelligible voice. In such cases, they were encouraged to continue talking or to speak up. As was mentioned earlier, one student (LS) did not manage to finish her third task within the one-hour time limit, and so the protocol she had generated on that task was discarded.

Phase 3: When students finished working on the tasks, they were asked to complete the follow-up questionnaire described earlier in the '*Data collection design*' section of the chapter.

The second phase of the session, when subjects were producing their verbal reports, was tape-recorded, while in the third phase, note-taking was, in most cases, sufficient for recording the subjects' comments on their ratings of the difficulty of the items. If, however, their comments were rather extended, the third phase, or relevant part of it, was also audio-taped.

5.3 Data analysis and results

5.3.1 Transcription of the protocols

5.3.1.1 Decisions that had to be made

In light of the amount of data collected, on the one hand, and the time-consuming nature of the process of data transcription and analysis, on the other, decisions had to be made with respect to ways of gathering adequate information from the data on the processes used and the difficulties met by the students responding to the tasks. It was decided to transcribe some of the protocols in full and use the transcripts for analysis in those cases, and analyse the content of the rest of the reports by listening to the audio-tapes. Transcribing at least some of the protocols in their entirety was considered important because the process of transcription helps the researcher understand in depth how the students went about processing the tasks and items, what exactly they were saying, how they arrived at their answers, whether correct or incorrect. It was also important that the set of protocols to be transcribed included two protocols from the same subject, i.e., two different tasks completed by the same subject, two protocols generated on the same

task, and at least one protocol on each task involved in the analysis. This resulted in the need to transcribe at least seven protocols. Ultimately, eight of the twelve verbal reports were transcribed, and four were analysed using the audio-tapes. The transcription of all eight verbal reports was carried out by the researcher, which provided sufficient experience in obtaining relevant information from the reports and an increased confidence for analysing the remaining four reports through repeated listening to the audio-tapes. (Sample transcripts and notes on students' task processing are given in Appendix C.)

5.3.1.2 The process of transcribing the protocols

Because the researcher lacked relevant skills and experience, transcription of verbal reports had to be learnt through doing it. Initially, the process included the following four main stages.

- *Listening to the report from beginning to end without stopping the tape*. The aim of this was to get an overall picture of the subject's processing of the particular task, check how detailed the report was, what sorts of difficulties to expect during the transcription process, and identify, if possible, parts of the report that were, for whatever reason, likely to be more difficult to transcribe than others.

- *Listening to the report as a whole for the second time*, in order to measure the time needed for the subject to complete the task, refine the overall picture of the subject's approach, identify and make notes on specific aspects of the way she responded to the task (e.g., whether she read the text or the options first, whether she read the whole text before trying to answer the items, whether she answered the items in the order in which they are presented in the task, etc.).

- *The actual transcription of the protocol.* This involved listening to the report in short sections of a few sentences or clauses and writing down every word used in those sentences. This process was rather long and tiresome. The source of difficulty in transcribing the exact words used by the subject is not necessarily related to the quality of the recording, although that can surely make transcription more difficult. The biggest difficulty results rather from the fact that, following from the nature of "thinking" aloud, the sentences to be transcribed are, more often than not, incomplete. The subject, whilst "thinking" aloud, may start a sentence but then, after a pause of a few seconds, may go on to refer to something else. For instance, a sentence started may reflect the subject's thinking about an answer to the last item on the task, but the words said aloud by her a second or two later may have nothing to do with the thought started, because she may already be thinking about a possible answer to an item, say, around the beginning or the middle of the text. It is not always easy for the transcriber to capture the exact words or phrases used if their meaning does not seem to fit in with the thought expressed in the utterance immediately preceding those words or they seem irrelevant at the given point in the subject's task processing. Besides, in attempting to respond to the items, the subject may frequently turn her attention from text to options and vice versa. However, from a single word or a phrase, even if the subject repeats it several times, it does not necessarily become immediately apparent whether she is trying to understand the meaning of a section of the text or one of the options. In general, the incomplete sentences, the fragmentary nature of utterances and the abrupt changes of thoughts and ideas in the report can make it rather difficult to follow how the subject proceeds with the task, which may, however, be crucial for the transcriber to be able to understand and transcribe the exact words used. Clarification of what is going on,

what exactly the subject is saying in a particular section of the report may require a certain amount of context, which, in turn, frequently involves the need for the transcriber to wind the tape backwards and forwards several times and, at the same time, keep consulting the student's task sheet. This requires much patience and time to be devoted to the task on the part of the researcher-transcriber. Typically, the more extended the report, the longer the process of its transcription.

- *Checking each transcript against the original recording,* in order to ensure that it indeed corresponds to the subject's verbalisation.

With experience gained from transcribing the first two or so protocols, the transcription process described above changed in a number of respects. First, it seemed helpful to include some more listening to the report as a whole before getting down to transcribing it section by section, utterance by utterance. The additional listening was to ensure the possibility of taking more detailed notes on how the subject proceeded in responding to the items before beginning the transcription. The notes related, in particular, to issues like how much of the text the student had read before attempting to answer an item, in what order she answered the items (including attempts to give answers), how many and which items she had already completed before answering an item, i.e., how many and which options she had to choose from when answering the item, which items were easy and which were difficult for her to complete, and the causes of difficulty in responding to individual items, as evidenced in the report. The detailed notes, requiring several listenings, made it much easier to follow the subject's processing of the task and, therefore, also to capture the exact words she had used in her report. In addition to this, the experience of transcribing the first few protocols resulted in some technical changes in the transcripts themselves. Specifically, it turned out to be helpful to segment the protocols, in order to make it apparent where an unfinished sentence ended and where a

new utterance or thought started, and also to employ time markers in the transcripts to indicate the duration of pauses, which would help the researcher infer aspects of the tasks that had caused difficulty for the students. Lastly, despite every effort to produce as complete and accurate transcriptions as possible, there remained occasional ambiguities in some of them, where it was not possible to clarify exactly what the subject was saying. Such ambiguities, together with unintelligible parts of the report, are also indicated in the transcripts. (Sample protocols are given in Appendix C: two from a 'strong' student [HS1] doing an easy task [T2 Giant Pandas] and a difficult task [T5 Caught out ..], and two from a 'weak' student [LMS] doing the same difficult task as the 'strong' student [T5 Caught out ..] and an easy task [T1 Julie wants ../Adverts].)

Each protocol was examined for its content, with a focus on students' overall approaches to texts and tasks, on the one hand, and the processes, skills and strategies they used when responding to individual items, on the other.

5.3.2 Analysis and results

The transcripts revealed that students had been able to think aloud as they carried out the tasks, and the frequency with which they verbalised activity descriptions was, in most cases, very low. The language of the report probably contributed to this to a great extent. Had the students been asked to generate their reports in English rather than in Hungarian, most protocols are very likely to have been much less informative than they are as a result of allowing students to use their mother tongue. The least detailed report, which is also the one where the frequency of verbalising activity descriptions is the highest, was produced by a lower level student (LMS) completing a difficult task (T5). Most probably, all students took a little longer to complete the tasks than they might

have done without the need to think aloud. The table below shows the time taken by students to complete each task, and the accuracy of their answers to the items.

Table 5.2 Time taken to complete each task and accuracy of answers

Task No	Task ID	k	S	Time	CA	S	Time	CA
T1	**Julie wants**	10	**LS**	18 mins	8	**LMS**	17 mins	8
T2	**Giant pandas**	5	**MS**	8 mins	3	**HS1**	12 mins	5
T3	**Gorillas in Uganda**	6	**LS**	29 mins	1	**LMS**	15 mins	4
T4	**Being wet …**	7	**MS**	34 mins	3	**HS1**	20 mins	7
T5	**Caught out …**	6	**LMS**	22 mins	1	**HS1**	35 mins	2
T6	**Animals under …**	8	**MHS**	23 mins	6	**HS2**	30 mins	5

Note: k=Number of items on the task; S=Student; CA=Number of Correct Answers

As can be seen from the table, the two lower level students (LS and LMS) took roughly the same time to complete Task 1 (18 and 17 minutes, respectively), and also gave the same number of correct answers to the items on the task (8 out of 10). However, the table also indicates that there must be considerable differences between these two students in terms of the skills and processes they used when completing another task, specifically, Task 3. LS needed almost twice as much time as LMS to carry out Task 3 (29 vs 15 minutes), and during that time she got only one out of six items right, in contrast to four correct answers in half the amount of time by LMS. Similar differences can be seen between MS and HS1, responding to Task 4. While MS took 34 minutes to complete the task, and answered three (out of seven) items correctly, HS1 got all seven items right in 20 minutes. The shortest time spent on a task was 8 minutes, which MS needed to carry out Task 2, even though she got only three (out of five) items right, while the longest time, 35 minutes, was taken by HS1 to answer the items on Task 5.

Details of how the students went about producing the protocols as well as similarities and differences in their approaches to completing the tasks will be discussed below.

Individual differences in the production of the protocols

LS provided detailed reports on processing her two tasks (Tasks 1 and 3). Apart from occasional requests to speak up when she was verbalising her thoughts in a very quiet voice, she did not need to be prompted to talk throughout the completion of the tasks. She had only very short pauses, typically, of 4-7 seconds during her reports. The protocols she generated clearly reflect her approach to text, as well as her test-taking behaviour, and provide useful information on her difficulties either in understanding the meaning of the texts she had to read or in responding to the items involved.

LMS produced protocols that were considerably different from one another in the amount of information included. She was able to produce the protocol as requested in the case of tasks that were easier for her to complete (Tasks 1 and 3). However, as briefly mentioned above, her report on processing a difficult task (T5) was the least extended of all the reports. During that task, she had long silences and, therefore, frequent prompts had to be used to encourage her to verbalise her thoughts. Although her protocol on the difficult task is much less detailed than her other two reports, it provides useful insights into the nature of her difficulties in answering the items on the task in question.

In contrast to LMS' verbal report on Task 5, HS1 generated a detailed and extended report on the same task, regardless of the fact that Task 5 was difficult for her as well. In fact, HS1 managed to think aloud successfully during the completion of all three tasks she had been assigned (Tasks 2, 4 and 5), irrespective of the difficulty of the task. She had relatively short pauses of, typically, no longer than 20-25 seconds. As a result,

all her protocols include very specific and detailed information on how she processed the tasks and how she arrived at an answer, whether correct or incorrect.

Similarly to HS1, MS, MHS and HS2 produced verbal reports that allow the researcher to follow the way they had proceeded with their tasks as well as to make inferences about their thought processes and the skills and knowledge they had used in their attempts to respond to individual items on the tasks. They did not need to be prompted during the production of their reports. In fact, MS talked continuously, with almost no pauses, while carrying out her tasks (Tasks 2 and 4). She even remarked that she liked completing tasks in this way, i.e., thinking aloud, because it was easier for her to complete a task if she said aloud everything she was thinking. All four students (HS1, MS, MHS and HS2) occasionally verbalised activity descriptions, which, however, they typically did in addition to, not instead of, verbalising their thoughts.

In sum, after training, all students managed to think aloud while completing their reading tasks and produced verbal reports that provide useful information on their approaches to the tasks, the skills and processes they had used and the difficulties they had encountered when answering the items. The only exception is the protocol that LMS generated on Task 5. It should be noted, however, that, in light of her other two protocols, LMS' failure to produce as complete and extended a report on Task 5 as was produced by HS1 appears to have resulted not so much from her (in)ability to think aloud as from the fact that the particular task, including the text involved, was simply too difficult for her.

Similarities and differences across individuals in overall approach to text and task

LS processed text very differently from all the other subjects. Typically, she either read the whole sentence and then translated it (or made attempts to do so), or, if the beginning of the sentence contained easy or familiar words, then, after reading, she translated those words, and only then went on to read the rest of the sentence. She generally read aloud those parts of the text, and reread those (parts of) sentences whose understanding caused her difficulty. She often repeated difficult or unknown words and phrases before either making a guess at their meaning or abandoning their translation. Occasionally, she tried to identify the main idea in a section of the text relying on her knowledge of easy, high-frequency words and ignoring unfamiliar words in the section, but more typically, she spent a lot of time trying to understand the meaning of every word in a given section. As became apparent from her protocols, she had many comprehension difficulties at both the more local level of the sentence and the level of the text as a whole. (Details will be given in the discussion of item level results.)

MS read some sections or sentences of the text aloud, others silently. Apart from reading aloud, she reread and translated, more or less, word by word, sentences of a section, or parts of a sentence, when she had difficulty in understanding either the main idea in the section or the meaning of the sentence as a whole. On the other hand, she summarized the content of those sections and sentences whose understanding did not require rereading on her part. She often ignored unfamiliar words and phrases, not even attempting to guess their meaning, when she (occasionally, wrongly) thought she had been able to identify the main idea in a given section without understanding the meaning of those words. In this respect, her approach is more like that used by the other four students, LMS, MHS, HS1 and HS2, who read the text, for the most part, silently,

section by section and attempted to understand it in as much detail as they thought was necessary in order to answer the questions/items. (As MHS reports, 'This / I didn't understand now but I don't find it that important.') After reading a sentence or two, these students typically paraphrased what they had read, or, if the section read was longer, they summarized the main points in the sentences involved or what they had understood from the information in those sentences, before going on to read the next section. Of the four subjects, the incidence of reading aloud or repeating parts of sentences is the highest in the case of MHS, which is likely to be due, to some extent, to the task type she completed (Task 6, '*Matching clauses to gaps in text*'). Specifically, while in the case of the other tasks, a correct answer generally required students to identify the topic and understand the main idea in different sections of the text, in Task 6, apart from understanding the main idea in a given section, the student had to understand the meaning of, occasionally, every word in certain sentences of the text, in particular, the gapped sentences to be able to select the correct answer. As is evidenced in MHS' protocol, reading aloud or repetition of words and phrases, in her case typically involved parts/clauses of the gapped sentences. (E.g., the clause involved in Item 1, '*Unless we act now …*', is repeated aloud four times at different points in her report.)

With respect to the students' approaches to the test task itself, LS again differs from all the other subjects. She read the text and task (questions/items) in the order in which they are arranged on the page. This was a suitable strategy for her to follow in the case of Task 1 ('*Matching statements to short texts*'), where the layout of the task is such that the questions precede the text (advertisements), but was a less successful strategy in the case of the other task (Task 3, '*Gorillas ..*' / '*Matching headings to text*'), where the text

comes first on the page, and the options (the section headings from which to choose the correct answers) are given after the text. In the case of Task 3, she spent 22 minutes reading and translating the text from beginning to end, before beginning to read the options (section headings) and, at the same time, answer the items, spending only 7 minutes total on the latter two activities. In answering the items, she followed the order of the options rather than the order of the items, that is, instead of searching for a suitable answer to each item (i.e., a suitable section heading for each section of the text), she tried to identify a suitable item for each option (i.e., a suitable section for each heading), as she was reading the list of section headings from top to bottom one by one.

The other five students (LMS, MS, MHS, HS1 and HS2) approached their tasks differently. They generally did not read through the whole text before reading the options and beginning to answer items on the text. They first looked through the options when, after reading the first or first two paragraphs of the text, they reached the Example Item (Item 0) or the first item to be completed. In this respect, the only exception is HS2, who first read through all the options, and only then started to read the text itself. However, all five students attempted to complete the items as they were processing the text section by section. The process included the following main steps. They

- read a section of the text that involved an item,

- looked through the options,

- answered or attempted to answer the item involved, and then

- went on reading the text up to the point where the next item was located.

They again looked through the options, answered or attempted to answer the item involved, and so on. The cycle was repeated until they reached the end of the text. In

this way, items/questions on a task that were easy for a student were answered during the first, section-by-section, reading of the text. Having read through the text once, answered as many items as they were able to, they returned to the unanswered items. This means that they reread shorter or longer sections of the text, as required by the items that were as yet to be completed, and tried to select suitable answers from the remaining (occasionally, much) smaller number of options.

While the five students' overall approach to the test task appears to be very similar, in some respects, there are notable differences in the way they processed their tasks. For instance, LMS generally read through the whole text in a relatively short time, regardless of the difficulty of either the text or the questions, because during the first reading, she did not spend much time on step 3) of the above cycle of task processing. That is, when reading the text for the first time, she either selected her answer to an item relatively soon, or left the item unanswered and went on reading the next section. In contrast, MS, completing Task 4 ('*Matching sentences to gaps in text*'), took rather long to get to the end of the text, as she spent considerable time trying to identify a suitable answer to the item involved in any given section of the text, before going on to read the next section. In addition, she did so after only glancing through the options, when more careful reading of the options would have been helpful if not crucial before making any attempts at answering the items. Even when, having read through the whole text, she decided, in her terms, to 'read' the options, she read them rather superficially. Unlike MS, HS1, completing the same task (Task 4), did a careful reading of the options when she reached the first gap in the text to be completed. As in this type of matching, crucial information is divided between the main text and the options (i.e., the sentences taken out of the text), from reading the options carefully, HS1 gained both a general idea of

the content of the text and some fairly specific information about the actual events of the narrative, before beginning to answer any item in the task. Apart from Task 4, Task 2 was also completed by both MS and HS1 and the strategy used by the latter student proved to have been more suitable in the case of both tasks.

Higher level students, in particular HS1 and MHS, appear to have been more systematic in their attempts to respond to the items than either LMS or MS, regardless of the type of matching they completed. Before finalising their answers to an item, HS1 and MHS systematically checked if the meaning of their selected option indeed fitted in with the meaning of the section of the text for which it was intended. Even when they had comprehension difficulties, they attempted to select their answers relying on what they had understood from the text and the options, rather than, as was more often the case with lower level students, on the basis of words which appeared both in a section of the text and in a given option. In the latter respect, HS2's approach to the items is similar to the approach used by HS1 and MHS. However, she was less systematic than the other two higher level students in checking the suitability of her answers. Apart from all this, there are some differences in the approaches and strategies used by the students that may result rather more from the nature of the task and text than individual differences between students, which will be briefly discussed in the section that follows.

Similarities and differences across texts and tasks

A 'word-matching' strategy was typically used in the case of Task 1 ('*Matching statements to short texts/advertisements*), where both students answered the majority of the items correctly by matching key words in the items/questions with the same or similar words in the text. In the case of all the other tasks, a correct answer to an item

more typically resulted from an understanding of the main idea in a given section of the text, on the one hand, and the meaning of the correct answer, on the other. However, as was hinted earlier, in Task 6 ('*Animals under threat*'/'*Matching clauses to gaps in text*'), knowledge of grammatical structures, an awareness of the structure of the gapped sentences of the text played an important part in getting items right.

While Tasks 2 ('*Giant pandas*') and 4 ('*Being wet ..*'), involving the same type of matching, specifically '*matching sentences to gaps in text*', appear to be similar in terms of the main approaches and processes that were used by students carrying out the tasks, they clearly differ in certain sub-skills and processes that were required for successful completion of the items on each. Due to the difference between the two texts, in the case of Task 4, in order to arrive at a correct answer, the student had to be able to process much longer and grammatically more complex sentences than in the case of Task 2, both as part of the main text and as options. In addition, in Task 2, students could answer most items if they understood the content of the particular section of the text in which an item was located and the meaning of the correct answer to that item, whereas in Task 4, to arrive at a correct answer, they often had to read beyond the section that involved the item they were about to complete, and understand some of the information that occurred before and/or after. As the protocols show, MS, as expected, was less successful in responding to the items on Task 4 than Task 2, and HS1 was more successful on both tasks than MS. (See Table 5.2)

In the case of Tasks 3 ('*Gorillas in ...*') and 5 ('*Caught out in the rain*'), both requiring '*matching headings to paragraphs of the text*', a failure to give a correct answer to an item often resulted from lack of knowledge of certain vocabulary items that would have been crucial for understanding the content of either a given section of the text or the

options (section headings). The protocols clearly demonstrate that the items on Task 5 were much more difficult for students than the items in Task 3.

Regardless of the nature of the text and the type of matching required, students not infrequently selected their answers, particularly when they had comprehension difficulties, by eliminating 'implausible' options first, and then comparing the content of the remaining options. If they failed to identify the answer to an item on the basis of what they had understood from the text and the options, as a last resort most students used the strategy of guessing an answer. Details of the skills, knowledge and processes students used when responding to each item on each task are discussed in the section that follows.

(When reading the following item by item account of the results of the analysis, the reader is advised to consult the relevant task at the same time. For the tasks, see Appendix A.)

Item level results

<u>**TASK 1**</u>: *Julie wants …*

Item type: Matching sentences to short texts

No of items: 10 items

Subjects: LS and LMS

Item 1

'*Jack wants something old and valuable.*'

Both LS and LMS got this item right. Both had difficulties in fully understanding the item/question, because of the unfamiliar word '*valuable*'. Both had to read all 16 options/advertisements and answer most other items on the task, before they were able

to complete this item. The process each followed in selecting the answer was, however, slightly different. LS considered two options as possible answers, the correct answer, on the one hand, and the only plausible distractor, in which, similarly to the correct answer, the word '*antique*' is used. Comparing the content of the two, she decided to choose the former, the correct answer, because she thought that '*antique clocks*' mentioned in that advertisement had probably more to do with the unknown word '*valuable*' in the question than '*antique style furniture*' mentioned in the other option. Similarly, for LMS, the same two options seemed plausible. However, as a result of her misunderstanding the key word in another item on the task (Item 8), she used, wrongly, the only plausible distractor as an answer to that misunderstood item. Thus, after eliminating all implausible options, for her there remained only one possible option to select, the correct answer.

Item 2

'*Jill wants a new pair of sandals.*'

Both students got the item right, easily recognising the match between the words '*sandals*' in the item/question, and '*shoes*' and '*feet*' in the correct answer.

Item 3

'*Angela wants to eat out with her boyfriend.*'

Both students answered the item correctly, again easily recognising the overlap between the very simple phrases '*eat out with her boyfriend*' in the item, and '*romantic meal*' in the correct answer.

Item 4

'*Charles wants to go on an exotic trip.*'

Both students got the item right. Although both of them failed to understand most of the information in the advertisement, they were successful in identifying the relationship

between the phrase '*go on an exotic trip*' in the question and the word '*Turkey*' in the advertisement. For LS, this was the first item on the task she responded to, because the correct answer to this item (Option A) comes first on the list of options and, in the process of answering the items, she generally followed the order of the options. LMS, on the other hand, answered the item only after she had answered seven other, easier, items on the task and had eliminated all implausible options. Even then she selected her answer with a degree of uncertainty, as is clear from her report. ('*In* [Item] *4, he'd like to travel . / well, one that may have to do with traveling might be the . the first one* [Option A] *at most because there it mentions Turkey . 6 persons'*.)

Item 5

'*Cathy has a toothache and wants a doctor.*'

LMS answered the item correctly, easily recognising the match between the words '*toothache*' in the item and '*dentist*' used in the advertisement, while LS failed to give any response to this item. LS had difficulties understanding the question itself, translating it as '*Cathy is a dentist and wants to be a doctor*', which may indicate problems with vocabulary, grammatical structure and/or superficial reading. Besides, however, she did not recognise the key word '*dentist*' in the correct answer, which might be due either to her unfamiliarity with the word, or to the use of difficult vocabulary in the immediate context of the word. The latter reason seems to be supported by the fact that she repeated the words preceding and following the key word several times before abandoning attempts to understand the information in the advertisement. ('*PUBA .. accredit dentist in the . Chorls . area . Chorles area / The only accredit . accredit dentist . . . accredit . . . /* only this BUPA *. . accredit . . . /* I don't know ..')

Item 6

'*Richard wants a band for his party*.'

Both students answered the item correctly, recognising the relationship between the phrase '*a band for his party*' in the question and '*Jazz .. music*' in the advertisement.

Item 7

'*Jane wants a new hairdo*.'

Both students got the item right, easily identifying the match between the words '*a new hairdo*' in the question and '*long hair*', '*salon*' in the advertisement.

Item 8

'*Peter wants home entertainment*.'

The item was difficult for both students. While LMS responded but gave an incorrect answer to the item, LS did not give any response. As is evidenced in their protocols, failure on this item was caused, in the case of both students, by lack of knowledge of the key word '*entertainment*' in the question. Although LMS gave an answer, she selected it only after she had completed 8 (out of 10) items on the task. She took the key phrase '*home entertainment*' in the question to mean '*home equipment*', and wrongly chose an advertisement used as a distractor that advertised '*furniture*'.

Item 9

'*Jessica wants new glasses*.'

Both students answered the item correctly, but with some difficulty. Neither of them knew one of the key words in the advertisement ('*spectacles*') and it took some time for both of them to recognise the other key word, used in the heading of the ad ('EYEWEAR'), before they were able to identify the match between the words '*glasses*' in the item, and '*eyewear*' in the correct answer.

Item 10

'*Roger wants pictures for his business.*'

LS answered the item correctly, recognising the semantic overlap between the question and the text, specifically, the phrase '*pictures for his business*' used in the question, and the word '*photography*' in the correct answer. Contrary to expectations, she did not consider as a possible answer the distractor which also uses the word '*photography*', because she understood the crucial information in that option. Unlike LS, LMS got this item wrong. Even before reading the option that provides the correct answer, she confidently selected her (wrong) answer on the basis of the relationship she had recognised between the word '*pictures*' in the question, and '*pencil*' and '*studio*' in an option also used as a distractor. Again surprisingly, she did not appear to consider choosing any of the options in which the word 'photography' appears. From her protocol, it appears that, in her mind, the word 'picture' is closely related to 'pencil' and 'drawing', while it has no meaning relationship with the word 'photography', used, among others, in the correct answer.

TASK 2: *Giant pandas*

Item type: Matching sentences to gaps in text

No of items: 5 items

Subjects: MS and HS1

Item 1

Both students answered the item correctly. They had no difficulties understanding either the main idea in the given section of the text or the meaning of the correct answer (i.e., the sentence that fitted the gap in the section). When choosing their answers, MS relied more on the exact word overlap between the paragraph and the correct answer than

HS1. Checking the suitability of her answer, MS referred to the fact that easy words like '*eat*', '*plant*' and '*bamboo*' appeared in both the paragraph and the option she intended to select as an answer. Although both of them read the preceding section, in addition to the section that included the gap, they identified the correct answer without reference to its content.

Item 2

Both students got the item right. Although MS had difficulties understanding certain details in the section, she understood the main idea, as well as the meaning of the correct answer, which was sufficient for her to answer the item correctly. As she reports, the relationship between the paragraph and her intended option '*is apparent because both present "data" about pandas*'. She answered the item during the first reading of the text. In contrast, HS1, despite (or because of) her apparent understanding of the content of the section in detail, skipped the item during her first reading of the text because, at that point in processing the task, she thought that more than one of the options/sentences might fit the gap in the section. As she says, ' *.. here after all . several . may fit so I'm going on . something will then get here on the basis of elimination.*' After responding to all the other items on the task, she selected the correct answer from the remaining (two/three) options very easily.

It is clear from the two protocols that the process of selecting the correct answer to this item was faster in the case of MS, who did not understand every detail in the section, and spent much less time thinking about the answer than HS1. It appears that a more detailed understanding of the text and the options became a disadvantage for HS1 when she had to answer this item.

Item 3

HS1 answered the item correctly, while MS got it wrong. HS1 had no difficulty understanding either the main idea in the section or the meaning of the correct answer. In addition to the two sentences preceding the gap, she also read the three-sentence long section that follows it, before selecting her answer. MS, however, understood the main idea in the paragraph only in part, because of her problems with the words '*area*' and '*diameter*' in the sentence immediately preceding the gap. More importantly, she did not understand the meaning of the correct answer either, apart from the meaning of the conjunction 'although' at the beginning of the sentence. Eventually, she chose an option (the distractor) whose content she had thought was related to the content of the section that comes right after the gap. However, this choice of answer was, in fact, based on the relationship she had identified between two short phrases, specifically, '*In the spring*' at the beginning of the sentence that immediately follows the gap, and '*In stormy weather*' at the beginning of the option she (wrongly) selected as an answer to the item. In her report, there is no evidence of her understanding the meaning of the two sentences she related to each other, apart from an understanding of the meaning of the above two phrases introducing those sentences.

Item 4

Both students got the item right, easily identifying the match between the particular section of the text and the correct answer. In addition to their understanding of both the main idea in the section and the meaning of the sentence that fits the gap in the section, both students recognised the lexical overlap between the phrase '*A new-born panda*', used in the text, and '*Pandas are born*', in the correct answer.

Item 5

HS1 responded to the item correctly, while MS got this item wrong. HS1 easily identified the relationship between the sentence that follows the gap and the correct answer, that is, the suitable sentence for the gap given among the options. Understanding the meaning of these two sentences was sufficient for her to give a correct answer to the item. It should be noted, however, that this is the last item on the task and by the time she came to answer this item, she had already answered all but one of the items on the task, therefore, instead of, say, six or seven options, there were only three from which she had to choose her answer.

MS also understood the sentence that follows the gap and, in her report, there is no sign of comprehension problems with the correct answer either. Her failure to identify the suitable answer may result from her superficial reading of the options, in general. Another reason might be related to the fact that the item is included in the very last, two-sentence long paragraph of the text, whose first sentence is the 'missing' sentence. From this fact she infers with no hesitation that the suitable sentence for the gap in the section must be the option beginning with the conjunction 'although', because, as she argues, '*They typically conclude a text with sentences like this: Although …*'. She selected her answer in accordance with this argument, not even considering the suitability of other options. While she is fairly confident of the adequacy of her (wrong) answer, it is clear from her protocol that, apart from the word '*although*', she understood very little of the content of the sentence she selected as an answer to the item.

TASK 3*: Gorillas in …*

Item type: Matching headings to paragraphs of a text

No of items: 6 items

Subjects: LS and LMS

Item 1

LMS got the item right, while LS got it wrong. LMS understood both the main idea in the target paragraph and the meaning of the option/paragraph heading considered to be the right answer (Option D, '*The location*'), and was able to recognise the match between the two. She answered the item after she had read the paragraph involved (second paragraph of the text/first numbered paragraph), without the need to read the rest of the text. LS, on the other hand, completed the item only after she had read the whole text. In her case, this was the only item on the task where, despite encountering a number of difficult, unfamiliar words, she was able to identify the main idea. She had no problem with the correct answer either, as she understood its meaning. Yet, unlike LMS, she failed to recognise the relationship between the two. One reason for this might be that reading through the whole text before attempting to answer the item, encountering even more difficult and unknown words during that process, and understanding even less from the information in the rest of the text, made her uncertain about the accuracy of what she had understood from the paragraph at the beginning of the text, even if the crucial information in the paragraph appeared to be relatively easy for her to understand while she was reading it. Experiencing considerable difficulties in understanding the text as a whole, when she began answering the items, she tried to identify matches between paragraphs of the text and the options, relying, for the most part, on overlapping vocabulary items. For this item, she (wrongly) selected the only option/paragraph heading in which the word '*population*', the last word of the target paragraph, is used.

Item 2

LS got the item wrong, while LMS answered it correctly. Although she understood some of the information in the paragraph and partially understood the meaning of the

correct answer as well (Option A, '*How the gorilla population is organised*'), LS wrongly selected her answer on the basis of the exact word match between the paragraph and an option in which the word '*group*', occurring twice in the paragraph, is used. LMS does not appear to have understood much more from the content of the paragraph than LS, which is likely to be the reason why she first thought to choose the same wrong option as LS. However, unlike LS, she did not have difficulties fully understanding the meaning of the correct answer and, therefore, when checking the content of the two superficially plausible options against the information she had understood from the paragraph, she was able to make the right choice. She responded to the item after reading the target paragraph, without the need to read the rest of the text.

Item 3

LS did not give any answer to this item, while LMS got the item right. LS did not even attempt an answer because, on the one hand, she failed to identify the main idea in the text and, on the other, lack of knowledge of the word '*leader*' prevented her from understanding the meaning of the correct answer (Option C, '*The leader of the group*'). LMS seems to have misunderstood the main idea in the paragraph, most probably because, similarly to LS, she was unfamiliar with the key phrase, '*led by*', in the section. She also had problems with the key word, '*leader*', in the correct answer, yet, unlike LS, she selected the correct paragraph heading. In fact, her correct answer is likely to have resulted from guessing both the main idea in the paragraph and the match between the content of the paragraph and the selected option. This seems to be supported by the fact that her explanation for her choice of answer is not very convincing and includes some contradiction: '*This* [paragraph] *is* [about] *[that] how many groups there are / I mean how many gorillas live and how and then that's* [Option] *C* ['The leader of the group'] *because groups are mentioned in that.*'

Item 4

Both students failed on this item. LMS did not give any response, but LS' wrong answer was, as is clear from her protocol, also only a guess. Their failure to identify the main idea in the two-sentence long paragraph appears to have resulted mainly from the fact that the paragraph contains many difficult words and phrases that were unknown to both of them (e.g., '*munched contentedly*', '*vegetation*', '*crashing down to provide*', '*sustenance*'). Similarly to the previous item, knowledge of the word '*leader*' would have been crucial for understanding the meaning of the correct answer to the item (Option H, '*What the leader of the group did*').

Item 5

LS did not answer the item, while LMS got it right. LS understood only a few easy, high-frequency words and phrases from the relatively long and difficult paragraph (e.g., '*wild animals*', '*modern world*', '*easy to lose*', *in the heart of Africa*', '*the life of these animals*'), which was insufficient for her to identify either the main idea in the paragraph or the relationship between the given section of the text and the right heading (Option E, '*Appreciation of a unique experience*'). The item was difficult for LMS. She only partially understood the main idea and, similarly to LS, she only knew one of the key words ('*unique*') in the correct answer. She responded to the item after she had read the whole text, answered four (out of six) items on the task and, thus, there remained only three options from which to choose the right heading for the paragraph. Although she understood much more of the information in the paragraph than LS, her choice of answer, as is clear from her protocol, was motivated, to a great extent, by her recognition of the connection between the words '*impressive wild animals*' in the text and '*unique*' in the correct answer.

Item 6

LS failed to understand the meaning of both the paragraph and the correct heading for it (Option G, '*What is done to protect the gorillas*'), yet she gave a correct answer to the item, because she found the same word, '*protect*', used in both. LMS also misunderstood (or did not understand) the main idea, and she did not (fully) understand the meaning of the correct answer, either. From her understanding of the general idea of the paragraph, '*It's about tourists and their attitude*', she inferred a relationship between the paragraph and an option (the distractor) that used the word '*reaction*' ('*The gorillas' reaction to seeing the author*'), and selected, wrongly, that option as an answer to the item. Unlike LS, she did not recognise the exact word lexical overlap between the given section of the text and the correct answer.

<u>**TASK 4**</u>: *Being wet*

Item type: Matching sentences to gaps in text

No of items: 7 items

Subjects: HS1 and MS

Item 1

HS1 answered the item correctly, while MS got it wrong. Apart from the paragraph that includes the gap (in fact, the first numbered gap in the text), HS1 also read the preceding three short paragraphs. She did not need to read any other sections of the text to be able to complete the item. She understood the main ideas in each paragraph she read, as well as the relationship across sentences of those sections, with no apparent problems with either vocabulary or grammar. She identified the correct answer (Option F, '*Eventually we wandered back to catch the 2 pm train home.*'), as soon as she had read it among the options. To make sure that the option she intended to select was indeed the correct answer, she read all the other options to check their content as well.

She also identified the distractor fairly quickly, on the basis of the information she had read in the first four paragraphs, on the one hand, and the sentences given as options, on the other. (As she reports, '*The motorway I think is a little further away from the railway station.*')

MS read the same four sections of the text as HS1 before her first (unsuccessful) attempts at answering the item. However, unlike HS1, she tried to identify a suitable sentence for the gap after only glancing through the options, relying, for the most part, on information expressed in easily understandable, high-frequency words and phrases in each. The main reason for her failure on this item appears to be the fact that she only understood the easier words in the correct answer ('*catch the 2pm train home*'), while she had difficulties understanding the meaning of the sentence as a whole. Specifically, due to lack of knowledge of the words '*eventually*' and '*wandered back*', she interpreted the correct answer '*Eventually we wandered back to catch the 2pm train*' as '*We caught the 2pm train*', which, thus, did not fit in with the meaning of the paragraph. Therefore she did not consider the (right) option to be a possible answer to the item, which resulted in her getting the item wrong.

Item 2

HS1 answered the item correctly, while MS got it wrong. As this item is included in the same section of the text as the previous item (the previous item at the beginning, this one at the end of the section), HS1 did not need to read any more of the text than she had done for the previous item to be able to select the correct answer. In fact, she answered this item, the second item on the task, before answering the previous item, i.e., the first item on the task, simply because the two answers are presented among the options (and therefore she read them) in this order. As she understood the meaning of the correct answer without difficulty (Option C, '*We were only caught in it for a minute*

but we were drenched – and we were only wearing flimsy T-shirts and sweatshirts.'), she easily recognised its relationship with the meaning of the sentences preceding the particular gap in the section.

MS identified the main idea in the text, but appears to have understood the correct answer only in part. It is clear that she understood, or found out by guessing, the meaning of the most important content words in the sentence in question (e.g., *'drenched', 'wearing', 'only', 'T-shirts', 'sweatshirts'*). However, in her report, there is no evidence of her understanding, or paying attention to, a grammatical item in the sentence, specifically, the reference word '*it*' (*'We were only caught in <u>it</u> for a minute but ..'*), which would have been crucial for a correct answer. In any case, she failed to recognise that the referent of the pronoun '*it*' used in the given option was the '*rain*' described in the sentences preceding the gap in the text she was trying to complete. Eventually she chose an option as an answer to the item whose meaning, due to partial understanding of its content, she had wrongly thought was related to the content of the section. The process of selecting the wrong answer is reflected in the following words of her report: '*The text before the gap talks about heavy rain and that they were caught in the rain, and [Option] A says that they were there without tea and everything. This seems to be a good match.*'

Item 3

HS1 answered the item correctly, whereas MS got it wrong. HS1 had difficulties understanding certain details in the section. Specifically, she had problems with the use of the substitution word '*one*' in the clause '*.. if it was the one going to our local station*'. This, however, did not prevent her from identifying the main idea in the section. She considered two options as possible answers to the item (Option A, and the correct answer, Option D). She chose the right answer by comparing the content of the

two options, checking the content of each against the content of the section and, from all this, making an inference with regard to which of the two was the more suitable sentence for the gap. Part of this process becomes apparent from her thoughts in the following short section of her report: '*Perhaps it's [Option] D that will be good after all because here . they don't want to argue / I can imagine as . they are drenched there . . / and maybe they still . have money . at the beginning / here in [Option] A it already writes that they don't / they don't have enough money even for a cup of tea*' [..] '*Then . for the time being I write in D*'.

MS, after several attempts, finalised her (wrong) answer when she had already answered five (out of seven) items on the task and was in a position to choose the answer for the two remaining items (this one and the next one) from only two options. Her response to both this item and the next one is a result of guessing, based mainly on the length of the two options rather than their content. Her reason for her final choice of answer is that, in her words, '*the longer sentence is more likely to come earlier in the text*'. The main source of her failure on this item is that she only understood isolated pieces information from the sentences preceding the gap in the section, which was insufficient for her to grasp the main idea of the section. Besides, she misunderstood the crucial information in the correct answer ('let the train leave'), taking the clause '*.. we just let the train leave the station*' to mean '*we left the station*'. Her comprehension problems with both the given section of the text and the suitable sentence for the gap in that section resulted in getting the item wrong.

Item 4

HS1 responded correctly, while MS got the item wrong. For HS1, the item became much more difficult than it would have been if she had not misread the word '*rain*' for '*train*' in the sentence coming immediately before the gap in the text. For quite a while,

she tried to make sense of the misread clause '*I couldn't believe they'd all avoided the (t)rain*', which was, however, an impossible enterprise because, in the given context, 'avoiding the train' made no sense. This confused her to the extent that she left the item unanswered during her first reading of the text. She returned to it after she had read the whole text, answered most other items on the task, and she had only three options to choose from. Despite her problem with the misread clause, she understood the main idea of the section, and had no difficulty understanding the meaning of the correct answer either, so, from the remaining three options, she confidently selected the suitable sentence for the gap.

MS was able to identify the main idea of the paragraph. However, she had considerable difficulties understanding the meaning of the correct answer (Option A, '*We'd have been happy to stand if they were worried we'd wreck the seat, but now we had to wait half an hour without even enough money for a cup of tea.*'). It appears that the main source of her problems was related to the length and complexity of the sentence in question, including the grammatical structures used in it. Her protocol shows that, in fact, she understood the meaning of only the very last clause of the complex sentence providing the correct answer, which is likely to be the reason why she did not recognise its relationship with the content of the paragraph. Her (wrong) response to the item, similarly to the previous item, is apparently a guess.

Item 5

Both students answered the item correctly. HS1 had no difficulty understanding the meaning of either the paragraph or the correct answer. Nevertheless, she responded to the item only after she had read the remaining two sections of the text and completed the two items involved in those sections. As her verbal report suggests, she found those two items slightly easier to answer than this item. MS also identified the main idea in

151

the paragraph. She did not fully understand the meaning of the correct answer, because she did not know, or remember, the word '*journey*', used in the phrase '*On the journey back home*' at the beginning of the sentence in question. However, she understood the crucial information in the sentence, ('.. *when I got back home I got straight into the bath* …') and was able to relate that information to the content of the paragraph.

Item 6

This was among the easiest items on the task for both students. They easily recognised the relationship between the text and the suitable sentence for the gap involved. MS' choice of answer was, admittedly, motivated, to some extent, by the lexical overlap between the item and the correct answer. (The word '*mum*' is used in both).

Item 7

Both students answered the item correctly. HS1 understood the meaning of both the sentence preceding the gap and the correct answer, and completed the item very easily. For MS, it took a little longer to select the answer, probably because she had difficulties in understanding earlier sections of the text. However, she realised that, as the item involves the very last sentence of the text, it required an answer that may serve as a concluding sentence for the text as a whole. As she fully understood the meaning of the correct answer and also understood the general idea of the text as a whole, she was able to select the suitable sentence for the gap.

TASK 5 *Caught out in the rain*

Item type: Matching headings to paragraphs of a text

No of items: 6 items

Subjects: LMS and HS1

Item 1

Both students got the item wrong. LMS first attempted to answer the item immediately after she had read the paragraph (the first item on the task), without reading the rest of the text. As she did not succeed in identifying a suitable heading, she decided to read the whole text before making further attempts to complete the item. The apparent reason for her failure on the item is that, on the one hand, the paragraph was too difficult for her to identify the main idea in it and, on the other, because of her unfamiliarity with the word '*short-cut*', she did not understand the meaning of the correct answer (Option G, '*Possible short-cut*'), either. Eventually, she (wrongly) selected an option (Option F '*A sudden obstacle*') that has a lexical overlap with the paragraph. (The second word of the first sentence in the paragraph is '*suddenly*'.) Her uncertainty during the process of making a decision about the item is clearly reflected in the following words from her report: '*Well that* [paragraph] *talks about something sudden and accidental and that perhaps matches* [Option] *F.*'

HS1 also had to read the whole text before she was able to select an answer. After initial problems with the word '*short-cut*' in the correct answer she found out its meaning, so her failure on this item is probably due to the fact that she did not fully understand the main idea of the paragraph, and the reason for this appears to be twofold. First, she had difficulties understanding certain words in the paragraph ('*glancing through*', '*rear entrance*', '*lobby*') whose knowledge might have been helpful for her to identify the main idea with more confidence. A more important reason than this, however, is that, although she skimmed through the preceding section of the text, which provides the Example (marked with 0), she did not read it very carefully and did not pay attention to the details of crucial information in that paragraph, which would have been necessary for her to be able to fully understand the main idea in the paragraph of this item. All this

left her with one plausible option, the distractor, which she eventually wrongly selected as an answer to the item.

Item 2

LMS failed to respond to the item and HS1 got it wrong. LMS understood some of the information in the paragraph, which might have been sufficient for a correct answer. However, as she was unfamiliar with the word '*obstacle*' used in the correct answer (Option F, '*A sudden obstacle*'), she failed to recognise the relationship of the option in question with the paragraph and used it (wrongly) as an answer to another item on the task.

HS1 clearly understood both the main idea in the paragraph and the meaning of the correct answer, yet she got the item wrong. In fact, at one stage in the process of responding to the task, immediately after she had read the paragraph, she gave a correct answer to the item. Her verbal report shows that her first choice of answer was not a result of guessing, but was based on her recognition of the relationship between what she had understood from the paragraph and the meaning of the option (paragraph heading) considered as the correct answer. The main reason why later she changed the right answer to the wrong one is the result of her misunderstanding the main idea of another paragraph of the text (Paragraph 4/Item 4). As, despite all her efforts, she was unable to identify a suitable heading for that paragraph (Paragraph 4) from the options she had not used earlier, she reconsidered her answer rightly given earlier to this item and (wrongly) decided to use it as an answer to that item (i.e., Item 4). This meant that ultimately she got both items wrong. It is apparent from her report that there is much uncertainty behind her final decision on the answer to both items involved.

Item 3

LMS did not respond to the item and HS1 got it wrong. Unfortunately, LMS' report provides no useful data on the sorts of problems she had specifically with this item. Her general comment on the difficulty of the text as a whole partially explains why: '*The trouble is I don't understand the text.*'

HS1 understood the main idea in the paragraph, yet she got the item wrong. The apparent reason for this is that she misread the word '*trick*' for '*brick*' in the correct answer (Option A, '*A trick – will it fail?*'). Misreading the key word in the correct answer meant that the option in question, losing its originally intended meaning, could not be related to the content of not only the paragraph whose understanding is in focus of this item, but to any other paragraphs of the text, either. As a result, HS1 eliminated the option in question as a possible answer to any items on the task, which led to total confusion when she tried to decide which of the remaining options/headings might be suitable for which paragraph of the text. As her protocol shows, misreading the crucial word in the answer to this item contributed to her failure not only on this item, but also on two other items (one preceding, the other following this item).

Item 4

Despite claiming that she did not understand the text (see quote above), LMS answered the item correctly, whereas HS1 got it wrong. LMS understood the main idea of the paragraph, and knew the key word ('*escape*') in the correct answer (Option B, '*An unexpected narrow escape*'), which enabled her to select the right heading for the paragraph. HS1 got the item wrong, principally because she misunderstood the main idea in the paragraph. In her protocol, there is no obvious sign of her either understanding or not, or even paying attention to the first, easily understandable, sentence of the paragraph, which carries crucial information for a correct answer ('*I was*

saved by the bell.'). Besides, unlike LMS, she seemed to have difficulties with the phrase '*be off*', whose understanding was, again, helpful for LMS in identifying the main idea in the paragraph. Lastly, as mentioned earlier, the confusion that resulted from misreading the key word in the correct answer to the previous item also appears to have contributed to HS1's failure on this item.

Item 5

LMS got the item wrong but HS1 answered it correctly. LMS was unable to identify the main idea in the paragraph. She also had difficulties understanding the meaning of the correct answer (Option C, '*Two approaches to public use of office buildings*'), as she was unfamiliar with the word '*approaches*' and could not find out the meaning of the phrase '*public use*', even though she knew both words of the phrase. She selected her (wrong) answer on the basis of overlapping vocabulary (the word '*housed*' is used in the paragraph, '*home*' in her selected option/heading). Unlike LMS, HS1 identified the main idea of the paragraph, despite her apparent difficulties understanding the text in detail. Although she was also unsure about the accuracy of her understanding of the word '*approaches*' used in the correct answer, this did not prevent her from selecting the option as an answer to the item immediately after she had read the paragraph.

Item 6

LMS failed to respond to the item, whereas HS1 answered it correctly. HS1 easily identified the relationship between the paragraph and the correct answer on the basis of her knowledge of the phrase '*Get off my land*' used in the paragraph, and the saying '*An Englishman's home is his castle*' in the correct answer. LMS was unfamiliar with both, which resulted in her failure on the item.

TASK 6 *Animals under ..*

Item type: Matching clauses to gaps in text

No of items: 8 items

Subjects: MHS and HS2

Item 1

Both MHS and HS2 answered the item correctly, but with considerable difficulty. Both read the three sentences of the text that precede the gapped sentence, and understood the main idea in those sentences. HS2 understood also the meaning of the gapped sentence itself ('*Unless we act now __* '), whereas MHS had difficulties understanding that, probably because of the conjunction 'unless'. Neither student had difficulty understanding the necessary information in the correct answer (Option E, '*in 50 years' time elephants and rhino will inhabit only the echoing corridors of museums or the territory of a zoo*'), although both appeared to be unfamiliar with the word '*inhabit*' used in it. Although HS2 understood everything that was, in principle, necessary for a correct answer, similarly to MHS she was unable to answer the item when reading the text for the first time, and had to return to it after she had read through the whole text. The two students arrived at the correct answer through different processes.

MHS first returned to the item when she had answered four (out of eight) items on the task and had, instead of the initial nine options, only five from which to choose the answer. At that point, she again made an attempt to respond to the item, but still could not make her decision. She then identified the answer, albeit with some uncertainty, when she had answered two more items, which meant that there remained only two items for her to complete (this item and Item 3), and she was in a position to choose the answer from only three options. Although she had some vocabulary problems with two

of the remaining three options, she understood their content in sufficient detail to be able to make the right choice.

HS2, in fact, identified the relationship between the item and the correct answer (Option E) immediately after she had read the gapped sentence ('*Unless we act now*'). However, as this was the first item on the task, she wanted to make sure that the option she (rightly) intended to select as an answer to the item was not suitable to any other item on the task and, therefore, she decided to go on reading the text without writing her intended answer in the answer box. However, when she read the next sentence of the text, which was gapped for the next item on the task (Item 2, '*the fact is that*'), she got confused because she thought that the option she (rightly) intended to choose as an answer to this item (Option E, the correct answer) might be a suitable answer to the next item (Item 2) as well. At that point, she looked through all the options again and eliminated those that she thought could not be suitable answers to this item for grammatical reasons. During that process, she identified the correct answer also to the next item (Option C for Item 2), but she had the same problem with the intended option (Option C) as earlier with the option she intended to select as an answer to this item, specifically, that it might be a suitable answer not only to the next item (Item 2) but also to this item (Item 1).

Despite her understanding of both the content and the grammatical structures involved in the two options she was considering (Options E and C), when she came to finalise her answer to either this item (Item 1) or the next item (Item 2), she was unable to decide which option to choose for which item. When comparing the two options, she did not observe any difference between their content because she thought they 'e*xpress the same idea in different words, one saying that species / animals will die out, the other*

that in 50 years' time they will only live in museums or a zoo.' It took her quite a long time to realise that, grammatically, only one of the two options was a suitable answer to the item. The process of making her final choice and her reasoning is reflected in the following sentence from her protocol: '*We can't say unless we act now, 1,000 species die out every year, we can only say what will happen in the future if we don't act now.*'

Item 2

Both students answered the item correctly. Both recognised the relationship between the item and the correct answer after they had read the gapped sentence ('*His view is not exaggerated or alarmist: the fact is that __*') and, in the case of MHS, the sentence immediately following the item. Similarly to the previous item, the process through which they arrived at the right answer was different. HS2, as described above, identified the answer to the item, but then she thought that the intended option might be a suitable answer to another item as well (specifically, to the previous item, Item 1). In fact, of the two items in question, first she finalized her answer to the previous item (Item 1) and then for her there remained only one plausible option regarding this item, the correct answer (Option C, '*around 1,000 of our bird and animal species become extinct every year*'). MHS, after reading the gapped sentence, first thought of choosing an "incorrect" option (Option J) as an answer to the item. As she was rather unsure about the adequacy of her choice, she decided to read on. On reading the sentence that comes immediately after the gapped sentence, she returned to the options, checked their content again, abandoned her earlier (wrong) choice (Option J) and fairly confidently selected the correct answer on the basis of its information content, before moving on to the next item.

Item 3

Both students got the item wrong. In both cases, the apparent reason for failure on this item is that they had difficulties in understanding both the gapped sentence ('*Do we want to preserve tigers, for example, ___ pacing up and down in a zoo?*') and the meaning of the correct answer (Option I, '*just so that our grandchildren can gape at them*'). With respect to the gapped sentence, the phrase '*preserve tigers*' caused difficulty for them, even though, after a while, HS2 guessed that it meant something like 'protect tigers'. Perhaps more important than this is that neither of them were familiar with the crucial phrase '*pacing* up and down in a zoo', following the gap in the sentence. Both students were unfamiliar with two key phrases used in the answer to the item, one of a lexical type ('*gape at*'), the other, which appears to be the more important of the two, grammatical ('*just so that*'). Although HS2 again guessed, from the context of the phrase '*gape at*', that it was very likely to mean 'see' ('*our grandchildren can* [see] *them*'), she took the adverb '*just so that*' to mean 'just like' expressing comparison. Partially as a result of this, she (wrongly) used the answer to this item as an answer to another item, which meant that it was impossible for her to identify the right answer to this item, however hard she tried. After several attempts, she finalised her (wrong) answer when she had completed all the other items and had only two options to choose from. Of the two, she selected the one that would have fitted the gap in the sentence, both grammatically and in terms of its content, if the missing clause had been the end, and not the middle part, of the gapped sentence. The reason why she seems to have ignored what came after the gap in the sentence might well be related to the fact that, as mentioned earlier, the gap is followed by a phrase ('*pacing up and down*') that was unknown to her. Similarly to HS2, this was the last item that MHS answered. Of the two options remaining after she had (rightly or wrongly) responded to all the other

items on the task, MHS chose the one that had an 'exact word' lexical overlap with the item. ('*preserve tigers*' is used in the first clause of the gapped sentence in the text, while '*preserve ... animals in their natural habitats*' is used in the option she wrongly selected as an answer to the item). She admitted that her answer to this item was a guess.

Item 4

Both students gave a correct answer to the item. They understood the meaning of the gapped sentence, including the simple but crucial phrase '*In a world __*' immediately preceding the gap, and had no difficulty understanding the meaning of the correct answer, either (Option A, '*where wonders of wildlife are available at the flick of a television switch*'). Nevertheless, neither of them was able to identify the answer when reading the text for the first time. From MHS' protocol, it appears that one reason for this might be related to the fact that the answer is included as the first item on the list of options (Option A). As she had only used one of the options before she first attempted to answer the item and, thus, there were still many options whose suitability had to be checked, she seems to have simply overlooked the first option on the list, beginning the process of checking with the second option (Option B). In any case, she recognised the relationship between the gapped sentence and the correct answer only when she had once read the whole text, completed four items on the task, and returned to the as yet unanswered items. Likewise, HS2 answered the item when there remained only two items for her to complete, including this one, and three options from which to choose the answer. At that point, however, she selected the correct answer very easily and quickly. Her approach to the item, the skill/knowledge and/or strategy she used for identifying the answer can be inferred from the following words in her report: 'I should

have noticed earlier that [Option] A is the good one for [Item] 4, because grammatically "*In a world*" [in the text] and then "*where ..*" [in Option A]..'

Item 5

MHS got the item wrong, while HS2 answered it correctly, albeit not without difficulty. From MHS' report, it is unclear whether or not she understood the meaning of the gapped sentence ('*So wouldn't it be better __ ?* '). It is, however, apparent that she had comprehension problems with the answer to the item (Option H, '*to pour the time and money into preserving these animals in their natural habitats*'), including vocabulary problems. HS2 clearly understood the gapped sentence but, similarly to MHS, had difficulties understanding the meaning of the correct answer. This is likely to be one of the reasons why, although eventually she got the item right, she did not respond to it very easily. Another thing that appears to have made the item difficult for her is that first she tried to identify a suitable answer mainly on the basis of the grammatical structure used in the gapped sentence. She thought that the option that might fit the beginning of the sentence '*So wouldn't it be better ..*' should start with '*if they ..*' or something like this, and there was no such option provided. Partially as a result, at one point in the process of completing the items she (wrongly) used the answer to this item as an answer to another item on the task, which, until she realised her mistake, prevented her from being able to identify a suitable answer to this item. She answered the item correctly when she had only two options to choose from. At that point, she easily recognised that, of the two, grammatically only one was a suitable answer (the one that used the 'to' infinitive).

Item 6

MHS answered the item correctly, while HS2 got it wrong. Although both of them identified the topic of the short section of the text that preceded the gapped sentence,

they appear to have understood only some details of the information in the particular section. They had no problems with the correct answer (Option B, '*because of their genetic make-up*') but, because of unknown vocabulary (e.g., '*clutch of*', '*hatchlings*' and, in the case of MHS, also '*tortoise*'), they only partially understood the meaning of the gapped sentence itself ('*In one clutch of eggs from a giant tortoise, for example, there will be some hatchlings which, __ , will develop longer necks than others.*'). Although MHS answered the item correctly, she did not easily recognise the relationship between the gapped sentence and its missing part, i.e., the correct answer. She skipped the item when reading the text for the first time and returned to it several times during the process of completing the task. She responded to it only when, having already answered five (out of eight) items on the task, she was in a position to choose the answer from only four options. HS2 considered two options as possible answers to the item, the correct answer (Option B) and the distractor (Option F, '*if there are practical reasons*'). Similarly to MHS, she also checked the meaning of each against the meaning of the gapped sentence but, after some time spent thinking over the suitability of the two options, unlike MHS she found the meaning of the distractor more suitable. Although she was not sure about the adequacy of her final choice of answer, she tried to reason that the resulting idea in the gapped sentence, specifically '*if there are practical reasons, they [animals] will develop longer necks*', seemed to be in accordance with the theory of evolution, mentioned in the given section of the text.

Item 7

MHS got the item right, while HS2 got it wrong. HS2 understood the crucial information in the gapped sentence ('*In times of drought, they will be able to reach the higher leaves __ , and so survive.* '), as well the meaning of the correct answer (Option D, '*which haven't yet been eaten*'), yet she was unable to recognise the

relationship between the two. From her report, no straightforward reasons can be inferred with respect to why this was so. What is clear is that she selected her (wrong) answer to the item fairly early in the process of completing the items on the task, when the number of options to consider was still rather high. As it is always more difficult to select an answer from a higher number of options, it may be the case that the relatively short clause involving the correct answer simply escaped her attention when she was searching the list of options for a suitable answer. Clearly, this is only speculation about something that is impossible to observe. MHS, on the other hand, got the item right during her first reading of the text, despite misunderstanding crucial information in the gapped sentence, confusing '*higher leaves*' with 'higher levels'. It appears that she answered the item correctly by identifying the relationship among the phrases '*be able to reach higher*', used before the gap, '*not eaten*', in the correct answer, and '*and so survive*', at the end of the gapped sentence. However, her idea of the meaning of the completed sentence, specifically that the particular unknown animal is able to 'get higher', and therefore, 'it is not eaten', and so can survive, is completely different from the meaning of the sentence in question.

Item 8

Both students answered the item correctly while reading the text for the first time. HS2 understood both the content of the target section, including the gapped sentence itself ('*In fact, of all the animals which have lived on earth, __* '), and the meaning of the correct answer (Option J, '*95% have either evolved into something else or have become extinct*'), without difficulty. MHS also understood most of the information in the section, had no problems with the gapped sentence either, but had some difficulties in fully understanding the meaning of the correct answer, mainly because of the unknown word '*evolved*'. Nevertheless, similarly to HS2, she recognised the connection between

the content of the gapped sentence and the information she had understood from the correct answer.

5.4 Summary of the results

In this chapter we have examined characteristics of the tasks and items under investigation from the perspective of the test taker. We have employed verbal protocol analysis to explore what cognitive processes, skills and strategies are actually used, what difficulties are encountered by students when responding to these tasks and items. Each task and item was completed by two students, specifically, a lower and a higher level student, and the protocols generated were analysed.

The analysis has shown that there are similarities as well as great differences in both students' overall approach to text and task, and the skills, processes and strategies they used when responding to individual items. Of the six students involved in the study, one (LS) processed both text and task very differently from the others. She generally read and attempted to translate the text word by word, while all the other students, including the other two lower level students (LMS and MS), typically processed the text section by section, tried to understand it in only as much detail as they thought was necessary to answer the items on the text, and generally paraphrased or summarised what they had understood. On the other hand, it has been observed that, in the case of all six students, the occurrence of re-reading and/or reading aloud parts of the text (words, phrases, and sentences) was triggered, for the most part, by a comprehension failure. The only exception to this was HS1, who also read aloud words and idiomatic expressions that she liked in the text.

In terms of task processing, the analysis has revealed that, contrary to expectations, students generally did not read through the whole text before beginning to answer items on the text. With the exception of LS, they all attempted to respond to the items during their first reading of the text. This approach was found to be more successful in the case of some students (e.g., HS1) than others (e.g., MS), even when completing the same task (T4), and more suitable in the case of some tasks (e.g., T3) than others (e.g., T5), regardless of the type of matching required (Tasks 3 and 5 involve the same type of matching – matching headings to text). The most important differences across students in their approach to processing the tasks and items appear to be related to the issue of 1) how much time they spent trying to identify a match between the content of a given section of the text and the suitable answer to the item involved before going on to read the next section of the text, 2) how carefully they read either the text or the options when the selection of the correct answer required careful reading, 3) how systematic they were in checking the suitability of their intended or selected answers, 4) whether they were able to make intelligent guesses at the meaning of crucial but unfamiliar words and phrases in either the text or the options from which to choose the answers, 5) whether they were able to eliminate incorrect options that had a semantic overlap with the correct answer, and, lastly, 6) the extent to which they relied on the lexical overlap between the item (a given section of the text) and the correct answer (or incorrect options, for that matter), when selecting their answers.

The item level analysis has also shown that it is not always the case that students answer an item correctly when they understand the content (main idea) of the relevant section of the text, that is, when they demonstrate the skill or knowledge necessary to answer an item (which Green, 1998, calls 'false negatives'), while there are cases when they get

the item right despite an apparent comprehension failure ('false positives'). Such cases were identified in the protocols of most students, irrespective of their language level (E.g., LS: Task 3 Item 1 – 'false negative'; LS: Task 3 Item 6 – 'false positive'; LMS: Task 3 Item 5 – 'false positive'; HS1: Task 5 Item 2 – 'false negative'; HS1: Task 5 item 3 – 'false negative'; MHS: Task 6 Item 7 – 'false positive'; HS2: Task 6 Item 7 – 'false negative')

Finally, the analysis has shown that the two students completing the same tasks and items very often used different processes, skills and knowledge, whether they eventually got the item right or wrong.

5.5 Concluding remarks

The verbal report data has provided us with significant insights into the actual processes that students went through when responding to individual items on the reading tasks under investigation. Clearly, such information on the tasks and items helps us better understand what is actually involved in taking reading tests of the kind and, therefore, findings of the study might be considered useful not only from a theoretical perspective but also from the perspective of test developers, whether they are designing tests for the classroom or a high-stake language examination. The next chapter will examine whether and to what degree findings of our analysis of verbal protocols are in agreement with findings of the content analysis of the items, and if there is a relationship between the item characteristics identified and the empirical difficulty of the items.

Chapter 6 Study Three: Exploring relationships among data sources

6.1 Introduction

The study in Chapter 4 described the content of the tasks and items focused on in this research and identified item characteristics believed to affect performance on these items. The study in Chapter 5 then, using verbal protocols, explored the actual processes students had used in producing answers to the items. The current chapter is intended to focus on the relationships among different types of information on the items obtained from the three main data sources involved in the research: content analysis, think-aloud protocols, and the empirical estimates of the difficulty of the items. It has two main aims. The first is to investigate the relationship between content analysis (Study One) and VPA (Study Two) and explore the value of using content analysis in specifying the skills, knowledge and processes required for successful completion of such reading items. The second is to discover more about possible reasons for differences in the difficulty of these items, relating findings of Study One and Study Two to the empirical difficulty of the items. The two main questions this chapter aims to answer are the following:

» *Do students process these reading items in the way predicted by content analysis?*

» *How does the information gathered on the items through content analysis and VPA relate to the difficulty of the items?*

Both questions, however, are very general and, therefore, the following more specific research questions were formulated:

Research Question 3: *Is it possible to observe in students' verbal reports the item characteristics identified through content analysis? In other words, do the verbal protocols provide evidence of the use of the skills, knowledge and processes that were predicted to be involved in responding to the items? If yes, to what extent do the two sets of data on the items agree?*

Research Question 4: *Is there a relationship between (any of) the item characteristic variables identified and the difficulty of the items? If yes, which item characteristics can be observed to have an impact on the difficulty of the items, and what are their specific effects? In other words, which (if any) of the item characteristics identified prove to make the items easier or more difficult to answer?*

The methodology used to seek answers to the above research questions (**RQ3** and **RQ4**) is described in the section that follows.

6.2 Methodology

To answer the question related to the agreement between Content Analysis and VPA (**Research Question 3**), first, it was necessary to find an appropriate procedure for comparing data from Study One and Study Two, bearing in mind that both studies are by their nature primarily qualitative ones. Eventually, it was decided to examine the issue in two different ways. One approach, which will be referred to in the discussion of the results as **Method 1**, involved a comparison of the descriptions of the content of the items (from Study One) with the think-aloud protocols (from Study Two), with a focus on identifying evidence in the verbal protocols for the **main process**, **skill** or **difficulty** that, on the basis of the item descriptions, appeared to best characterize each item.

However, as it has been observed in both Study One and Study Two that there are typically multiple processes going on in answering these reading items, it was found useful to explore the match between the two sets of data also from the perspective of **each individual item characteristic variable** identified in Study One. The results of the latter comparison will be discussed under **Method 2**.

The actual procedure in the case of **Method 1** was to inspect each item description and the relevant (parts of the) verbal protocols, and indicate against each item whether there was evidence in the protocol of each test taker (by ticking Yes or No) for the predicted *main process or difficulty*. The ratio of the number of identified cases of evidence in relation to the total number of possible cases across the items and the verbal reports was taken as a broad indicator of the degree to which the test takers involved in the study had used the predicted processes. As this procedure was, in important respects, based on the researcher's own description of the items, her familiarity with finer details of the items (as well as of the verbal protocols) might have affected/biased her judgement of the presence or absence of evidence in the protocols. To control for the effect of potential bias in assessing evidence for the main process, an independent judge, an applied linguist with expertise in assessing reading, in test construction and evaluation and in language test research was asked to carry out the above task of matching the content analysis data to the verbal reports. He was provided with copies of the relevant documents (the item descriptions, on the one hand, and students' verbal reports and notes on the reports, on the other), and on completing the task, he emailed the results to the researcher.

The procedure in the case of **Method 2** involved, first, tagging segments of each verbal protocol with appropriate variable codes, which, in effect, meant applying to the VPA data the coding framework developed and used to code the items in the content analysis study (Study One). Variable occurrences coded for each item in the protocol of each test taker, that is, *observed* variable occurrences, were then compared to the *predicted* occurrences of *each variable* across the items.

Applying to the VPA data the framework of 22 item characteristic variables used in the content analysis study was considered important not only for the reason mentioned above but also for two further reasons. Firstly, it enabled the researcher to make efforts at quantifying the verbal report data, similarly to the way it was done in the content analysis study, in a more finely tuned manner, which, it was thought, would help to lend more rigour to statements about the degree of agreement between these two sets of data on the items and, ultimately, between findings of Study One and Study Two. Secondly, it was found useful also for our investigation into the issues of item difficulty raised by **RQ 4**, since the resulting data made it possible to examine effects of the item characteristic variables on the difficulty of the items from the perspective of both "predicted" and "observed" variables.

To explore the issue of whether any of the 22 variables identified related to the difficulty of the items (**Research Question 4**), the data obtained on *predicted* and *observed* occurrences of each individual variable were analysed in three steps, in three different ways, each involving a different type of analysis. The intention of **the first type** of analysis was to find out if certain variables occurred (in either set of data) markedly **more frequently** with '**difficult**' items than with '**easy**' ones or vice versa.

The basic assumption underlying this approach was that difficult items were likely to share at least some common features missing from easy items and, therefore, if, in our data set, a variable occurred in all or most of the difficult items but in none or only a few of the easy ones or vice versa, then that could be considered as a piece of evidence supporting the particular variable's contribution to the difficulty (or 'easiness') of the items involved. To carry out this analysis, first, it was necessary to determine, on the one hand, the groups of "difficult" vs "easy" items and, on the other, the groups (or sets) of items associated with each variable. The former was done in the following way. The 42 items involved in the study were rank-ordered from the most difficult one to the easiest according to their empirical estimates of difficulty and then the list of rank-ordered items was divided into two halves, with an equal number of 21 items in each. The items in the top half, with their IRT logit values ranging between 3.06 and -0.51, were classified as "difficult", while those in the bottom half, with logit scores ranging between -3.26 and -0.52, as "easy". To determine the sets of items associated with each variable, all coding data were entered in a Q-matrix format, which is an incidence matrix capable of displaying the relationship between the items and the variables. The matrix consists of the number of variables by the number of items, with ones and zeroes indicating the occurrence of each variable by item. If an item involves a variable, then that is coded as 1, otherwise as 0. The bottom row of the matrix shows the totals for each column/variable (Hatch and Lazaraton 1991; Buck et al. 1997; Yong-Won Lee and Yasuyo Sawaki 2007; Hae-Jin Kim, Yasuyo Sawaki and Claudia Gentile 2007). (See Q-matrix A - CA and Q-matrix B - VPA in Appendix D). In the case of the verbal report data, an item was coded on a variable if the variable in question occurred in at least one of the two students' verbal protocols. Once the data were prepared for analysis, frequencies of variable occurrence in what came to be called the 'Top' (difficult) vs

'Bottom' (easy) Groups of items were compared in terms of both Content Analysis and VPA data.

The second type of data analysis involved an examination of **variable-based item difficulties**. The purpose of the analysis here was to see whether the difficulty level of the sets of items associated with each variable reflected different/distinguishable levels of difficulty, in other words, whether there was a hierarchical relationship among the variables in terms of the difficulty level. For this purpose, the average item difficulty for each variable was calculated, using the IRT estimates of item difficulty, and the data obtained in this way was analysed.

In **the third approach** to exploring effects of the item characteristics identified on item difficulty, the scope of the investigation was broadened to include in the analysis also the data obtained from the follow-up questionnaires completed by each subject at the end of the think-aloud sessions (see Study Two), and the issue was explored from the perspective of **students' perception** of the difficulty of the items. Ratings from the two students completing the same items were examined in relation to each other, the accuracy of the answers, the variables observed in each student's verbal protocol, and the item difficulty estimates.

Throughout this study, for numerical analyses, including mainly descriptive statistics, correlations between relevant sets of the data and data displays, the computer programme Microsoft EXCEL was used. Further details about the methods employed will be given in the section presenting and discussing the results, to which we shall turn now.

6.3 Results and discussion

6.3.1 Relationship between Content Analysis and VPA (RQ3)

6.3.1.1 Method 1: Comparing CA and VPA data with a focus on main processes

Table 6.1 presents the results of the analysis that aimed to examine if there was **evidence** in students' verbal protocols **for the main process** or **difficulty predicted** for each item. The table displays the results by item across the six tasks for each of the two students who completed the same tasks. The accuracy of students' answers to the items is also indicated to assist the interpretation of the results.

Table 6.1 Comparison of item descriptions with verbal protocols

TASK 1 *Julie Wants*

Item Number	Student 1 LS	Response	Student 2 LMS	Response
1	Y	1	Y	1
2	Y	1	Y	1
3	Y	1	Y	1
4	Y	1	Y	1
5	0	9	Y	1
6	Y	1	Y	1
7	Y	1	Y	1
8	Y	9	0	x
9	Y	1	Y	1
10	Y	1	0	x
TOTAL	9/10	8/10	8/10	8/10

Codes: Y=Yes, there is evidence 0= no evidence; 1=correct 9=blank x=wrong

TASK 2 *Giant Panda*

Item Number	Student 1 MS	Response	Student 2 HS1	Response
11	Y	1	Y	1
12	Y	1	Y	1
13	0	x	Y	1
14	Y	1	Y	1
15	0	x	Y	1
TOTAL	3/5	3/5	5/5	5/5

Codes: Y=Yes, there is evidence 0= no evidence; 1=correct 9=blank x=wrong

TASK 3   *Gorillas*

Item Number	Student 1 LMS	Response	Student 2 LS	Response
16	Y	1	0	x
17	Y	1	Y	x
18	0	1	0	9
19	Y	9	0	x
20	0	1	0	9
21	0	x	0	1
TOTAL	3/6	4/6	1/6	1/6

Codes: Y=Yes, there is evidence 0= no evidence; 1=correct 9=blank x=wrong

TASK 4   *Being Wet …*

Item Number	Student 1 MS	Response	Student 2 HS1	Response
22	Y	x	Y	1
23	0	x	0	1
24	0	x	0	1
25	0	x	0	1
26	Y	1	Y	1
27	Y	1	Y	1
28	Y	1	Y	1
TOTAL	4/7	3/7	4/7	7/7

Codes: Y=Yes, there is evidence 0= no evidence; 1=correct 9=blank x=wrong

TASK 5   *Caught out in the rain*

Item Number	Student 1 LMS	Response	Student 2 HS1	Response
29	0	x	0	x
30	0	9	Y	x
31	0	9	0	x
32	Y	1	0	x
33	0	x	Y	1
34	0	9	Y	1
TOTAL	1/6	1/6	3/6	2/6

Codes: Y=Yes, there is evidence 0= no evidence; 1=correct 9=blank x=wrong

TASK 6   *Animals under threat*

Item Number	Student 1 MHS	Response	Student 2 HS2	Response
35	0	1	Y	1
36	0	1	Y	1
37	0	x	Y	x
38	0	1	Y	1
39	0	x	Y	1
40	0	1	Y	x
41	0	1	0	x
42	0	1	0	1
TOTAL	0/8	6/8	6/8	5/8

Codes: Y=Yes, there is evidence 0= no evidence; 1=correct 9=blank x=wrong

A frequency count of the identified cases of evidence shows that there was evidence in the verbal protocols for the predicted main process in 47 items / cases out of the total number of 84 possible cases across items and individuals (2x42 items). This indicates a 56% agreement between the two sets of data.

On closer examination of the table, we can see that, in agreement with some of the observations in Study Two, individuals varied considerably in the degree to which their verbal protocols showed evidence of the predicted main process. Partially as a result of this, there are also considerable differences in evidence available in the protocols across individuals on the same task. For instance, if we look at the results for Task 1, it can be seen that the predicted process was evidenced for all 10 items on the task and, in the case of 7 of these items, evidence was found in the protocol of both students completing the task, whereas for the items in Task 6, all six cases of evidence available were identified in one of the two students' protocols. If we look also at Task 5, the results show that the predicted process was evidenced in the case of four out of six items on the task, however, each of the four cases were identified in the protocol of either one or the other of the two students completing the items. These observations are summarised in Tables 6.2 and 6.3 below.

Table 6.2 Evidence available by student

Student	Evidence available	%
LS	10 / 16 items	62
LMS	12 / 22 items	54
MS	7 / 12 items	58
MHS	0 / 8 items	0
HS1	12 / 18 items	66
HS2	6 / 8 items	75

Table 6.3 Evidence in common across individuals on the same task

Task	No of items on task	No of items with the predicted process evidenced	%	Evidence in common across individuals	%
1	10	10	100	7	70
2	5	5	100	3	60
3	6	3	50	1	16
4	7	4	57	4	57
5	6	4	66	0	0
6	8	6	75	0	0

From a different perspective, Table 6.1 also reveals that evidence of the predicted main process was identified in the case of a number of items where the accuracy of students' answers indicates either a failure to answer the item, or a failure to answer it correctly (e.g., Task 1 Item 8 by Student 1 / LS; Task 5 Item 30 by Student 2 / HS1), which appears to support relevant findings in Study Two, suggesting that successful completion of the items, in many cases, involves factors other than the main process or ability that the item is intended to measure. The investigation of the effects of such factors forms a part of the analyses in later sections of this chapter.

Lastly, when interpreting the results of the analysis presented above and assessing the degree of agreement between the two sets of data, it is important to consider two points related to the methodology, that is, Method 1, used here. First, there might be various reasons why a protocol does not show evidence of the main process or ability expected to be involved in an item. Some of these might be the following.

1   The item requires the use of a process or ability other than what was predicted. (The prediction was wrong).

2   The student did not possess the predicted process or ability and, therefore, she could not use it in answering the item. (As a result, she either failed on the item or got the item right using, for instance, the strategy of guessing.) (The prediction might be correct.)

3   The student's problems with unfamiliar/difficult vocabulary in relevant sections of the text prevented her from being able to use the predicted main skill or process (e.g., identifying the main idea). (In which case, the prediction concerning the main skill or process might be correct; the prediction of the involvement of processing difficult vocabulary is correct and evidenced.)

4   The student did use the predicted skill or process in responding to the item but this may not be apparent from her verbal protocol, for instance, because the transcript of the protocol is not sufficiently complete or because she did not report using the skill. (The prediction might be correct.)

Second, the task of matching the two sets of data with a focus on the predicted main process may prove to be a rather difficult and time-consuming activity for the rater, as was reported to be the case by the expert carrying out this analysis. The reason for this might be two-fold. For one thing, because of the interaction between various processes involved in responding to an item, when carrying out the activity, it may often turn out to be difficult for the rater to decide which of the predicted processes to consider as the main process used in any particular case by the student in attempting to answer the item and, technically, which predicted process to tally as evidenced in the verbal protocol. Point 3 above may illustrate the case, where it may not be easy to decide whether to tally the *lack of evidence* for the ability to identify the main idea, or tally the *presence of evidence* for processing difficult vocabulary. For another thing, however complete and accurate a transcript may be, without the rater being present during the think-aloud session, it is likely to be more difficult for her/him to follow students' processing of the items and, consequently, to infer the subjects' thought processes from what may often seem to be unrelated words and sentences in the transcript of a protocol.

Considering some of the limitations on the method used to compare the two sets of data in the first step of our investigation, the second approach to examining the same issue may also serve to validate the results discussed above.

6.3.1.2 Method 2: Comparing CA and VPA data in terms of individual variables

Table 6.4 below presents the results of coding the VPA data for the 22 item characteristic variables identified in Study One (CA Study). To enable comparison, a separate column (Predicted Variables) is included in the table to show the results of coding from Study One. The accuracy of students' answers, also indicated, may help understand the occurrence of one or another variable in a given item. Variable codes will be explained as necessary at relevant points in the discussion of the results. In addition, an extract from the coding framework used to code the items in the content analysis study is provided below. For the framework with detailed descriptions of the variables, see Study One.

Key to variable codes:

Text-related variables
v1, v2, v3, v4 → variables related to linguistic characteristics of the text
v5 → a variable related to the topic of the text

Item-related variables
v6, v7, v8 → variables related to item-type
(v6 – locating specific information; v7 – understanding main idea; v8 – understanding structural relations within sentence)
v9, v10, v11, v12 → variables related to the language of the question/correct answer

Variables related to the scope of the relationship between text and item
v13, v14, v15 → variables related to the amount of processing required by the item
v16, v17, v18, v19 → variables related to lexical overlap
v20, v21 → variables related to the elimination of superficially plausible incorrect options
v22 → a variable related to the elimination of syntactically inappropriate options in the case of matching-clauses-to-gaps-in-text type of items

Table 6.4 Distribution of observed variables across items and individuals

Item No	Predicted Variables (Content Analysis)	Observed Variables (VPA)	AC	Observed Variables (VPA)	AC
		Student 1 (LS)		**Student 2 (LMS)**	
1	v6, v10, v13, v17	v6, v10, v17	1	v6, v10, v21	1
2	v6, v13, v16	v6, v16	1	v6, v16	1
3	v6, v13, v16	v6, v16	1	v6, v16	1
4	v6, v13, v16	v6, v16	1	v6, v16	1
5	v6, v13, v16	v10	9	v6, v16	1
6	v6, v13, v16	v6	1	v6, v16	1
7	v6, v13, v16	v6, v16	1	v6, v16	1
8	v6, v10, v13, v18	v10	9	v10	0
9	v6, v10, v13, v16	v6, v10, v16	1	v6, v10, v16	1
10	v6, v10, v13, v17	v6, v16 , v17	1	-	0
		Student 1 (HS1)		**Student 2 (MS)**	
11	v7, v14, v16	v7, v14, v20	1	v7, v14, v16	1
12	v7, v14, v16	v7, v15, v20	1	v7, v14, v16	1
13	v7, v10, v11, v14, v16	v7, v14	1	v10	0
14	v7, v14, v18	v7, v14, v16	1	v7, v14, v16	1
15	v7, v15, v16	v7, v14	1	v7	0
		Student 1 (LS)		**Student 2 (LMS)**	
16	v7, v14, v19	v7	0	v7, v14	1
17	v1, v3, v7, v12, v14, v19	v7, v10, v12	0	v7, v19	1
18	v1, v3, v7, v10, v14, v17	v10, v12	9	v10, v12	1
19	v4, v7, v10, v15, v20	v10, v12	0	v10	9
20	v3, v4, v5, v7, v10, v14, v16	v10, v12	9	v16	1
21	v3, v4, v5, v7, v12, v14, v16	v10, v12, v16	1	-	0
		Student 1 (HS1)		**Student 2 (MS)**	
22	v1, v7, v10, v15, v17, v21	v7, v15, v20	1	v10, v12	0
23	v1, v7, v9, v11, v12, v14, v17, v21	v7, v15, v20	1	v7	0
24	v1, v2, v7, v10, v12, v15, v18, v21	v7, v15, v20	1	v12	0
25	v1, v3, v4, v7, v11, v12, v15, v21	v7, v12, v15, v21	1	v7	0
26	v1, v2, v7, v11, v15, v18	v7, v15, v21	1	v7, v15	1
27	v1, v2, v7, v11, v14, v16	v7, v15	1	v7, v15, v16	1
28	v7, v15, v21	v7, v15	1	v15	1
		Student 1 (HS1)		**Student 2 (LMS)**	
29	v1, v2, v4, v7, v9, v15, v19, v21	v9, v19	0	v9	0
30	v1, v2, v4, v7, v9, v15, v21	v7, v14	0	v9	9
31	v1, v7, v10, v15, v16, v21	v7, v10	0	-	9
32	v3, v7, v9, v15, v17, v21	v9	0	v7, v14	1
33	v1, v2, v3, v4, v5, v7, v10, v14, v16	v7, v10, v14,	1	v10	0
34	v4, v5, v7, v14, v17	v7, v14	1	v10	9
		Student 1 (MHS)		**Student 2 (HS2)**	
35	v2, v5, v8, v10, v11, v15, v20, v22	v10, v15	1	v8, v15, v20, v22	1
36	v2, v5, v8, v9, v15, v20, v22	v8, v15, v20	1	v8, v15, v20, v22	1
37	v2, v4, v5, v8, v9, v14, v19	v4, v9, v10, v15	0	v4, v10	0
38	v2, v5, v8, v14, v22	v8, v15	1	v8, v15, v22	1
39	v2, v5, v8, v9, v14, v22	v15	0	v8, v10, v15, v22	1
40	v4, v5, v8, v9, v15, v21	v4, v10, v15	1	v20, v22	0
41	v5, v8, v12, v15, v21	v7, v15, v20	1	-	0
42	v1, v2, v5, v8, v9, v14, v17, v22	v7, v14	1	v14	1

AC=Answer Correct; 1=correct answer 0=wrong answer 9=blank

An examination of Table 6.4 reveals several points, which are summarized below.

First, as expected, the number of variables predicted to be involved in any particular item is typically higher (and, in many cases, much higher) than what can be observed for each item in the verbal protocols.

Second, seven of the 22 item characteristic variables identified in Study One do not occur in any of the verbal reports. Four of these are related to *characteristics of the text* (**v1** – target paragraph uses syntactically complex sentences, **v2** – sentences of the target paragraph tend to be long, **v3** – sentences of the target paragraph uses the passive voice, and **v5** – topic of the paragraph is abstract), the fifth relates to *the length of the correct answer* (**v11**), the sixth to the amount and *way of text processing (scanning)* required by the item (**v13**), while the seventh is a lexical overlap variable (**v18**). Note that the first six of these seven variables involve item characteristics that, while representing important characteristics of reading items in general, as they were operationalised, defined and used to code the items in our Study One, can hardly be expected to occur in think-aloud protocols.

Third, typically, the variables occurring in the verbal protocol of the student who answered an item correctly is not observed in that of the other student who failed on the same item. For instance, in the case of **Item 5**, Table 6.4 reveals that neither of the two predicted variables (v6 and v16) that occur in the verbal protocol of Student 2 (LMS), who got the item right, occurs in Student 1's protocol (LS), who failed to give an answer to this item. For another example, in the case of **Item 22**, of the three variables (v7, v15, and v20) that can be observed in the verbal report of Student 1 (HS1), who answered the item correctly, none occurs in the other student's verbal report (Student 2 / MS), who got the item wrong.

Fourth, in cases of failure on the item, the verbal reports, more often than not, show the lack of occurrence of predicted variables related to

• *item type*

(v6 – locating specific information, v7 – understanding main idea, and v8 – understanding structural relations within sentence), which might be worth considering when interpreting the results of the analysis that focused on the main process (Method 1 discussed above) (see e.g., **Item 5** by LS, **Item 13** by MS, **Items 8, 19, 29**, and **37** by both students, **Item 39** by MHS),

• *the process of eliminating incorrect options*

(v20, v21, and v22) (e.g., **Items 23** and **25** by MS), and

• *the lexical overlap between the item and the correct answer*

(v16) (e.g., **Item 20** by LS, **Item 21** by LMS).

Fifth, looking at the above cases from the opposite direction, the variables that do occur in the verbal reports in cases of failure on the item are often associated with 1) difficulties in understanding *key vocabulary in the question and / or the correct answer* (v9 and v10) (e.g., **Item 5** by LS, **Item 8** by both students, **Item 32** by HS1) and 2) difficulties with *grammatical structures used in the question and / or the correct answer* (v12).

Sixth, in cases of successful completion of the item, the verbal reports generally show certain predicted variables to occur in the items in combination. **Items 23** and **25** may illustrate the point. HS1 gave a correct answer to both items, while MS got both items wrong. If we look at the variables occurring in their verbal protocols, we can see that, in the case of both items, in HS1's verbal protocol, v7 (the ability to understand main idea)

occurs along with one or another of the variables related to the elimination of incorrect options (v20 and v21). In contrast, MS' verbal report shows the occurrence of only v7 in the case of both items. All this suggests that, in addition to understanding the main idea of the passages involved, MS, similarly to HS1, should have also been able to eliminate incorrect options to be able to answer these items correctly.

Seventh, the observed occurrences of v20 and v21 appear to highlight the difficulty involved in specifying whether, in any particular case of responding to matching items, the process of eliminating incorrect options will involve making inferences on the basis of information in the target paragraph only (v20), or test takers will also rely on what they have understood from various other sections of the text (v21).

Last, it is also clear from Table 6.4 that the verbal reports not infrequently show the occurrence of certain variables in items in which they were not predicted to occur.

For the purpose of quantitative analyses of the relationship between predicted and observed variables, the frequency of observed occurrences of each variable across the items was counted. However, as a simple count of variable occurrence across the items would have distorted the results at item level, which was also considered to be an issue of interest, it was necessary to determine also the frequencies with which the variables were observed to occur in the particular items in which they had been predicted to occur. This required, first, the identification of the items observed to involve each variable, and then checking the identified cases against those that had been predicted to involve each variable. As mentioned earlier, each time an item was observed in the verbal protocols to involve a variable was considered to represent one case even if the particular variable occurred in the verbal report of both students completing the same item. The data obtained in this way made it possible to compare predicted and observed

variable frequencies at the level of not only the set of reading items examined but also at the level of individual items. Table 6.5 displays the observed vs predicted cases of occurrence of each of the 22 item characteristic variables identified in Study One, whilst Table 6.6 presents the results of the quantitative analyses.

Table 6.5 Items associated with each variable

Variable	Item Number	
	Predicted	*Observed*
v1	17, 18, 22, 23, 24, 25, 26, 27, 29, 30, 31, 33, 42	-
v2	24, 26, 27, 29, 30, 33, 35, 36, 37, 38, 39, 42	-
v3	17, 18, 20, 21, 25, 32, 33	-
v4	19, 20, 21, 25, 29, 30, 33, 34, 37,40	37,40
v5	20, 21, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42	-
v6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	1, 2, 3, 4, 5, 6, 7, 9, 10
v7	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34	11, 12, 13, 14, 15, 16, 17, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 41, 42
v8	35, 36, 37, 38, 39, 40, 41, 42	35, 36, 38, 39
v9	23, 29, 30, 32, 36, 37, 39, 40, 42	29, 30, 32, 37
v10	1, 8, 9, 10, 13, 18, 19, 20, 22, 24, 31, 33, 35	1, 5, 8, 9, 13, 17, 18, 19, 20, 21, 22, 31, 33, 34, 35, 37, 39, 40
v11	13, 23, 25, 26, 27, 35	-
v12	17, 21, 23, 24, 25, 41	17, 18, 19, 20, 21, 22, 24, 25
v13	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	-
v14	11, 12, 13, 14, 16, 17, 18, 20, 21, 23, 27, 33, 34, 37, 38, 39, 42	11, 12, 13, 14, *15*, 16, 30, 32, 33, 34, 42
v15	*15,* 19, 22, 24, 25, 26, 28, 29, 30, 31, 32, 35, 36, 40, 41	12, 22, 23, 24, 25, 26, 27, 28, 35, 36, 37, 38, 39, 40, 41
v16	2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 15, 20, 21, 27, 31, 33	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 20, 21, 27
v17	1, 10, 18, 22, 23, 32, 34, 42	1, 10
v18	8, 14, 24, 26	-
v19	16, 17, 29, 37	17, 29
v20	19, 35, 36	11, 12, 22, 23, 24, 35, 36, 40, 41
v21	22, 23, 24, 25, 28, 29, 30, 31, 32, 40, 41	1, 25, 26
v22	35, 36, 38, 39, 42	35, 36, 38, 39, 40

Table 6.6 Frequency of predicted vs observed occurrences of the variables

Variable	P	O		OAPI		UO		PNO
	n	n	%	n	%	n	%	n
v1	13	-	-	-	-	-	-	-
v2	12	-	-	-	-	-	-	-
v3	7	-	-	-	-	-	-	-
v4	10	2	20%	2	20%	0	0	8
v5	12	-	-	-	-	-	-	-
v6	10	9	90%	9	90%	0	0	1
v7	24	21	87%	19	79%	2	9%	5
v8	8	4	50%	4	50%	0	0	4
v9	9	4	44%	4	44%	0	0	5
v10	13	18	138%	11	84%	7	38%	2
v11	6	-	-	-	-	-	-	-
v12	6	8	133%	4	66%	4	50%	2
v13	10	-	-	-	-	-	-	-
v14	17	11	64%	8	47%	3	27%	9
v15	15	15	100%	9	60%	6	40%	6
v16	16	14	87%	12	75%	2	14%	4
v17	8	2	25%	2	25%	0	0	6
v18	4	0	0%	0	0%	0	0	4
v19	4	2	50%	2	50%	0	0	2
v20	3	9	300%	2	66%	7	77%	1
v21	11	3	27%	1	9%	2	66%	10
v22	5	5	100%	4	80%	1	20%	1
Tot.	**223**	**127**	**56%**	93/163	**57%**	34/127	**26%**	70 42%
	163	**127**	**77%**	93/127	**73%**			

Codes: **P**=Predicted ; **O**=Observed; **OAPI**=Observed As Predicted at Item level; **UO**=Unpredicted Observed; **PNO**=Predicted but Not Observed; *n*=number of cases;

As can be seen from Table 6.6, the total number of predicted cases of occurrence of the 22 item characteristic variables in the set of items examined is 223, while the total number of occurrences observed in the verbal protocols is 127, which indicates a 56% agreement between predicted and observed frequencies. If, however, we exclude from the analysis the six variables (see shaded cells of the table) that, as suggested earlier, can hardly be expected to occur in verbal report data, at least in the form they were

operationalised in Study One, the rate of agreement increases to 77%, which can be considered relatively high. The strength of the association between predicted and observed frequencies as determined by the Pearson correlation coefficient is r=0.61 when all 22 variables are included, and r=0.76 when the six variables in question are not included in the analysis.

When variable occurrence is examined at the level of individual items, again with the six variables excluded from the analysis, the results show that in 57% of the 163 predicted cases, the variables were observed to occur in the items in which they were expected to occur, while 42% of the predicted cases (70 cases) were not observed in the verbal protocols of the students participating in the study. Viewed from the perspective of observed occurrences, Table 6.6 shows that of the 127 observed cases, 73% were predicted, while 26% (34 cases) involve unpredicted occurrences of nine different variables. These nine variables are v7 (2 cases), v10 (7 cases), v12 (4 cases), v14 (3 cases), v15 (6 cases), v16 (2 cases), v20 (7 cases), v21 (2 cases) and v22 (1 case).

From a different perspective, Table 6.6 also reveals that the rate of agreement between predicted and observed frequencies of variable occurrence is much higher in the case of some variables than others. For instance, in the case of v4 (*There is lower-frequency vocabulary in the crucial information in the text*), it is 20%, which means that the verbal protocols showed evidence of the students' difficulties with vocabulary in the text in only 20% of the items predicted to involve this variable, while in the case of another vocabulary-related variable, specifically, v10 (*Key vocabulary in the question and/or the correct answer might be unfamiliar to lower-level students*), it is 138%. The latter figure indicates that the students participating in the study had problems with unfamiliar

vocabulary in the question and/or the correct answer in the case of a higher number of items than predicted (18 items instead of 13).

Of the three 'item-type' related variables (v6 – *locating specific information*, v7 – *understanding main idea*, and v8 – *understanding structural relations within sentence*), the degree of agreement is the lowest in the case of v8 (50%), while it is reasonably high, 90% and 87%, respectively, in the case of v6 and v7.

Looking at the four lexical overlap variables (v16, v17, v18, and v19), Table 6.6 shows that the agreement between predicted and observed frequencies is the highest (87%) in the case of v16 (*The item has lexical overlap with the correct option but not with the other options*), whereas, at the other end of the scale, v18 (*The item has lexical overlap with the correct option and one or more incorrect options but the overlap with the correct option is stronger than with the incorrect options*) was not observed in any of the verbal reports. One of the likely reasons for the lack of occurrence of v18 in the verbal protocols is that, in part, it overlaps with v16. However, it may also result from the relatively small sample of subjects providing verbal report data for this study.

The last point to make here concerns the predicted vs observed occurrences of v20 and v21 (both relating to the elimination of incorrect options). While v20 was predicted to occur in three items and was observed in nine, v21 was thought to characterise eleven items and was identified in the verbal reports in three cases. If, however, we look at the actual items associated with these two variables (see Table 6.5 'Items associated with each variable'), we can see that five of the nine items observed to involve v20, specifically, Items 22, 23, 24, 40 and 41, overlap with five of those eleven items that were predicted to involve v21. This reinforces our earlier tentative observation that, when coding the items, whether it involves content analysis or VPA data, it is often

difficult to make finer distinctions between various aspects of the process of eliminating incorrect options and, likewise, in term of the analysis in this study, between v20 and v21.

Implications for the next stage of the investigation

Results of the analysis discussed above raised issues for consideration with respect to our investigation into effects of the item characteristic variables identified on the difficulty of the items. In light of the results, it was felt useful, on the one hand, to merge some of the original 22 variables and, on the other, to discard those with no observed cases of occurrence in the verbal reports, unless the variable, despite the lack of its occurrence in the verbal report data, was considered to be a crucial characteristic of the reading items under investigation. Those variables among the original 22 that were either combined, or discarded from later phases of our investigation are shown in Table 6.7, while Table 6.8 presents the resultant 15 variables, along with the items associated with each after merging.

Table 6.7 Variables merged or discarded

Variables merged			*New variable code*
v1, v2, v3, v11	Linguistic characteristics of text/question or correct answer	→	v1
v4, v9	Lower-frequency vocabulary in text/question or correct answer	→	v4
v6, v13	Locating specific information/scanning	→	v6
v20, v21	Process of eliminating incorrect options	→	v20
v18	Discarded (lack of occurrence in verbal reports)		-

Table 6.8 Result of merging and discarding variables

Vari-able	Items associated with each variable			
	Predicted		Observed	
	Item number	n	Item number	n
v1	13, 17, 18, 22, 23, 24, 25, 26, 27, 29, 30, 31, 32, 33, 35, 36, 37, 38 39,42	**20**	-	**-**
v4	19, 20, 21, 23, 25, 29, 30, 32, 33, 34, 36, 37, 39, 40, 42	**15**	29, 30, 32, 37,40	**5**
v5	20, 21, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42	**12**	-	**-**
v6	1, 2, 3, 4, 5, 6, 7, 8, 9, 10	**10**	1, 2, 3, 4, 5, 6, 7, 9, 10	**9**
v7	11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28 29, 30, 31, 32, 33, 34	**24**	11, 12, 13, 14, 15, 16, 17, 22, 23, 24, 25, 26, 27, 28, 30, 31, 32, 33, 34, 41, 42	**21**
v8	35, 36, 37, 38, 39, 40, 41, 42	**8**	35, 36, 38, 39	**4**
v10	1, 8, 9, 10, 13, 18, 19, 20, 22, 24, 31, 33, 35	**13**	1, 5, 8, 9, 13, 17, 18, 19, 20, 21, 22, 31, 33, 34, 35, 37, 39, 40	**18**
v12	17, 21, 23, 24, 25, 41	**6**	17, 18, 19, 20, 21, 22, 24, 25	**8**
v14	11, 12, 13, 14, 16, 17, 18, 20, 21 23, 27, 33, 34, 37, 38, 39, 42	**17**	11, 12, 13, 14, 15, 16, 30, 32, 33, 34, 42	**11**
v15	15, 19, 22, 24, 25, 26, 28, 29, 30, 31, 32, 35, 36, 40, 41	**15**	12, 22, 23, 24, 25, 26, 27, 28, 35, 36, 37, 38, 39, 40, 41	**15**
v16	2, 3, 4, 5, 6, 7, 9, 11, 12, 13, 15, 20, 21, 27, 31, 33	**16**	2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 14, 20, 21, 27	**14**
v17	1, 10, 18, 22, 23, 32, 34, 42	**8**	1, 10	**2**
v19	16, 17, 29, 37	**4**	17, 29	**2**
v20	19, 22, 23, 24, 25, 28, 29, 30, 31, 32, 35, 36, 40, 41	**14**	1, 11, 12, 22, 23, 24, 25, 26, 35, 36, 40, 41	**12**
v22	35, 36, 38, 39, 42	**5**	35, 36, 38, 39, 40	**5**

The above 15 variables and the items associated with each, as shown in Table 6.8 above, formed the basis of the data set used in our examination of the relationship between item characteristic variables and item difficulty (**RQ 4**) in the next two sections of this chapter (Sections 6.3.2.1 and 6.3.2.2).

6.3.2 Relationship between the variables and item difficulty

6.3.2.1 Analysis of variable occurrence in 'difficult' vs 'easy' items

In the first step of our examination of the relationship between the variables identified and item difficulty, an analysis of the frequency of occurrence of each variable across 'difficult' vs 'easy' items was conducted to find out if any of the variables occurred markedly more frequently in either one or the other of the two groups of items. The distribution of variable occurrence across 'difficult' vs 'easy' items was expected to

give an indication of the impact of each variable on the difficulty of the items examined.

The analysis was carried out on data from both Study One (CA) and Study Two (VPA).

The results are presented in Tables 6.9 and 6.10 below.

Table 6.9 Variable frequencies in top vs bottom groups of items
(based on Content Analysis data)

Variable	Total	TOP GROUP ('difficult' items)		BOTTOM GROUP ('easy' items)	
	n	n	Ratio	n	Ratio
v1	20	16	80%	4	20%
v4	15	13	86%	2	13%
v5	12	10	83%	2	16%
v6	10	0	0%	10	100%
v7	24	13	54%	11	45%
v8	8	8	100%	0	0%
v10	13	6	46%	7	53%
v12	6	4	66%	2	33%
v14	17	7	41%	10	58%
v15	15	14	93%	1	6%
v16	16	2	12%	14	87%
v17	8	5	62%	3	37%
v19	4	2	50%	2	50%
v20	14	14	100%	0	0%
v22	5	5	100%	0	0%

Table 6.10 Variable frequencies in top vs bottom groups of items
(based on VPA data)

Variable	Total	TOP GROUP ('difficult' items)		BOTTOM GROUP ('easy' items)	
	n	n	Ratio	n	Ratio
v1	-	-	-	-	-
v4	5	5	100%	0	0%
v5	-	-	-	-	-
v6	9	0	0%	9	100%
v7	21	13	61%	8	38%
v8	4	4	100%	0	0%
v10	18	9	50%	9	50%
v12	8	4	50%	4	50%
v14	11	5	45%	6	54%
v15	15	13	86%	2	13%
v16	14	0	0%	14	100%
v17	2	0	0%	2	100%
v19	2	1	50%	1	50%
v20	12	9	75%	3	25%
v22	5	5	100%	0	0%

Nine of the 15 variables were found to have notable effects on the difficulty of answering the items involved in the examination (see shaded cells of Table 6.9). Seven of them appeared to make the items more difficult, whilst two were found to have 'easifying' effects.

The seven variables include three *text-related variables*, specifically, **v1** (long and syntactically complex of sentences in the target paragraph and/or the question and/or the correct answer, and the use of the passive voice in the text), **v4** (lower-frequency vocabulary in the crucial information in the target paragraph and/or the question and/or the correct answer), and **v5** (topic of the target paragraph is abstract), one item-type related variable, specifically, **v8** (*understanding structural relations within sentence*), one variable describing *the amount of text processing required* by the item, specifically, **v15** (a correct answer requires reading two or more consecutive sections of the text), and two variables associated with *the process of eliminating incorrect options* (**v20** – the elimination of incorrect options requires text-based inferences, and **v22** – the elimination of incorrect options requires syntactic knowledge).

One of the two variables that were observed to make the items easier include one item-type related variable, specifically, **v6** (*locating specific information* by scanning) and a lexical overlap variable, namely, **v16** (the item has *lexical overlap with the correct option but not with the incorrect options*).

As can be seen from Table 6.9, of the 20 occurrences of **v1**, 80% (16 cases) were identified in items in the Top Group ('difficult' items) and only 20% (4 cases) in items in the Bottom Group ('easy' items). The differences in ratio scores concerning variable frequencies across the two groups of items were even slightly greater than this in the case of **v4** (86% vs 16%), **v5** (83% vs 16%), **v15** (93% vs 9%) and, in the reverse

direction, **v16** (12% vs 87%), while in the case of the remaining four of the nine variables mentioned above, all occurrences were identified in either the Top Group (**v8**, **v20** and **v22**) or the Bottom Group of items (**v6**). Although of the latter four variables, for v22 (applying syntactic knowledge in eliminating incorrect options) a smaller number of occurrences were identified in our data set than for the others (only 5 cases), the fact that all of its occurrences are linked to items in the Top Group, that is, 'difficult' items appears to support its likely contribution to the difficulty of the items under investigation.

Regarding some of the other variables whose impact on the difficulty of the items might be worth noting, the rather even distribution of the occurrences of **v7** (understanding main idea) across 'difficult' vs 'easy' items (54% vs 45%) suggests that main idea items might be both difficult and easy to answer. A similarly even distribution in the case of **v14** (41% vs 57%), as opposed to **v15** (93% vs 6%), implies that items that require reading only one specific short section of the text (v14) might be both difficult and easy, while those that require reading and understanding information in longer sections of the text (v15) can be expected to be more frequently difficult than easy for students to answer.

Putting aside the findings related to v1 (linguistic characteristics) and v5 (topic of text is abstract), which variables might be difficult to identify in VPA data, and the fact that VPA data will, in general, show lower frequencies of variable occurrence than what may emerge from a content analysis of the items, the results of our analyses based on the two sources of data appeared to show the same tendencies with respect to effects of the variables on the difficulty of the items. With v1 and v5 excluded from the analysis, the Pearson correlation between 'predicted' and 'observed' frequencies of each variable

with respect to items in the Top Group ('difficult' items) is r=0.81, while the same indicator regarding the Bottom Group of items ('easy' items) is r=0.91.

Looking at the agreement at the level of individual variables, Tables 6.9 and 6.10 shows that, for instance, **v6** occurred in both sets of data in 'easy' items only (Bottom Group). In contrast, **v8** and **v22**, with each associated with fewer items than v6, though, were both predicted and observed to occur in 'difficult' items only (Top Group). **v4**, believed to characterize both 'difficult' and a few 'easy' items, was observed in fewer cases than predicted, all of its observed occurrences were, however, identified in the Top Group ('difficult' items). In contrast, **v16** and **v17**, which, likewise, were thought to characterize both 'difficult' and 'easy' items, were observed in the verbal protocols in 'easy' items only.

Lastly, it is also apparent from the tables above that, while **v10** (unfamiliar vocabulary in the question and/or the correct answer) showed in both sets of data an even distribution of occurrence across difficult vs easy items, **v20** (the process of eliminating incorrect options), being an equally important yet much less considered aspect of item difficulty in test development, appeared to characterize difficult items rather more than easy ones. As can be seen in Table 6.10 (VPA data), even the observed frequency of v20 is much higher in the Top Group (difficult items) (75%) than in the Bottom Group of items (25%), which indicates its potential impact on the difficulty of responding to reading items of the kind and, at the same time, highlights the importance of the variable for further research on item difficulty.

6.3.2.2 Analysis of variable-based item difficulty

In the second step of exploring effects of the variables identified, variable-based item difficulties were examined. The analysis was intended to find out about possible variations in the difficulty level of the sets of items associated with each variable. With respect to data from both Study One and Study Two, the average item difficulty for each variable was calculated, using IRT estimates of difficulty. Variable-based item difficulties are summarised in Tables 6.11 and 6.12.

Table 6.11 Variable-based item difficulty (CA)

Variable	*n*	*Minimum*	*Maximum*	*Average difficulty*	*SD*
v8	8	0.89	2.53	**1.77**	0.3
v22	5	0.89	1.96	**1.55**	0.3
v5	12	-1.28	2.53	**1.18**	0.8
v4	15	-1.28	3.06	**1.07**	1.0
v20	14	-0.51	3.06	**1.03**	1.0
v15	15	-0.56	3.06	**0.91**	1.0
v19	4	-1.28	3.06	**0.77**	2.0
v1	20	-1.31	3.06	**0.74**	1.0
v17	8	-1.32	1.66	**0.12**	0.9
v12	6	-1.28	1.93	**-0.05**	0.8
v10	13	-1.32	1.88	**-0.22**	0.7
v7	24	-2.63	3.06	**-0.23**	1.0
v14	17	-2.63	2.53	**-0.30**	1.2
v16	16	-3.26	1.44	**-1.44**	0.9
v6	10	-3.26	-0.69	**-1.90**	0.8

Table 6.12 Variable-based item difficulty (VPA)

Variable	n	Minimum	Maximum	Average difficulty	SD
v1	-	-	-	**-**	-
v5	-	-	-	**-**	-
v4	5	1.36	3.06	**2.23**	0.4
v22	5	0.89	1.98	**1.61**	0.3
v8	4	0.89	1.96	**1.52**	0.3
v19	2	-1.23	3.06	**0.91**	2.1
v15	15	-1.42	2.53	**0.65**	1.0
v20	12	-2.63	1.98	**0.28**	1.1
v10	18	-3.26	2.53	**-0.03**	1.2
v14	11	-2.63	2.26	**-0.07**	1.4
v7	21	-2.63	2.26	**-0.14**	1.0
v12	8	-1.28	0.51	**-0.53**	0.6
v17	2	-0.69	-1.32	**-1.00**	0.3
v16	14	-3.26	-0.93	**-1.93**	0.6
v6	9	-3.26	-0.69	**-2.02**	0.7

As can be seen from Tables 6.11 and 6.12, according to data from both studies, items associated with **v6** (locating specific information) and **v16** (the item has lexical overlap with the correct answer) were the easiest on average, and were occasionally much easier than items associated with the other variables. In contrast, items that required processing lower-frequency vocabulary in the text/the correct answer (**v4**) or abstract topic (**v5**), understanding structural relations within the sentence (**v8**) or applying syntactic knowledge in eliminating incorrect options (**v22**) were much more difficult than items not involving these variables. Of the latter four, in the case of v4, the average item difficulty based on verbal report data was considerably higher (M=2.23) than the difficulty level suggested by the data from Content Analysis (M=1.07). However, as indicated by estimated standard deviations, in the Content Analysis data, v4 showed a much wider range of item difficulties (SD=1.0) than in the VPA data (SD=0.4), which means that the students participating in the study did not have or, in any case, their verbal reports did not show evidence of having difficulties with lower-frequency

vocabulary in the case of many 'easier' items that had been predicted to involve such difficulties.

As regards the three item-type related variables, items that involved understanding structural relations within the sentence (**v8**) were much more difficult (M=1.77 / 1.52) than items that involved understanding main ideas (**v7**) (M=-0.23 / -0.14) which, in turn, as expected, were much more difficult than those that involved locating specific information (**v6**) (M=-1.90 / -2.02). Of these three, v7 showed the widest range of item difficulties (SD=1.0), while standard deviations indicated the least variability for v8 (SD=0.3) in both sets of data.

Items that required processing one specific short section of the text (**v14**) were relatively easier than those involving the ability to read and understand information across two or more consecutive sections of the text (**v15**). However, both variables showed a great deal of variability in item difficulties in both sets of data (SD=1.0-1.4).

Looking at the three variables associated with lexical overlap (**v16**, **v17**, and **v19**), it can be seen that items which had lexical overlap with the correct option but not with incorrect options (**v16**) were very easy on average (M=-1.44 / -1.93), whereas those that had lexical overlap with (an)incorrect option(s) but not with the correct option (**v19**) were relatively difficult in both sets of data (M=0.77 / 0.91). It should be noted that the frequency of occurrence of v19 was very low in our data set (4 cases in the CA data and 2 in the VPA data). However, despite its low frequency of occurrence in our data, its potential impact on the difficulty of the items appears to be supported by the fact that the difficulty level of those few items in which it occurred is considerably higher in both sets of data than the average difficulty of items associated with two of the three item-type related variables, specifically, v6 and v7. The same seems to apply to **v17** (the item

has lexical overlap with the correct option and one or more incorrect options), which, along with **v4**, **v8**, and **v22**, showed the least variability in the verbal report data.

It might be worth considering that for **v4** (lower-frequency vocabulary in the crucial information in the text / correct answer)**,** the data from Content Analysis showed much wider ranges of item difficulties than the VPA data (SD=1.0 vs 0.4), whereas for **v10** (key vocabulary in the question / correct answer might be unfamiliar to *lower-level students*), the verbal report data showed considerably more variability than the data from Content Analysis (SD=1.2 vs 0.7). The more important point is implied by the result concerning the latter variable (v10), suggesting that the students participating in the study had difficulties with unfamiliar vocabulary in the question and/or the correct answers in the case of a range of items where they had been supposed to be familiar with key vocabulary. As the Minimum values indicate, the easiest item for **v10** has a logit score of -1.32 in the Content Analysis data, while -3.26 in the verbal report data. The need to differentiate levels of vocabulary when describing the content of test items, as proposed by the framework of item characteristics in Study One, seems to be reinforced also by the striking difference in average item difficulty shown for **v4** (M=1.07 / 2.23), on the one hand, and **v10** (M=-0.22 / -0.03), on the other. Figures 6.1 and 6.2 below illustrate the hierarchical relationship among the variables as determined by the data from Study One, on the one hand, and Study Two, on the other.

Figure 6.1 Variable-based item difficulty (CA)
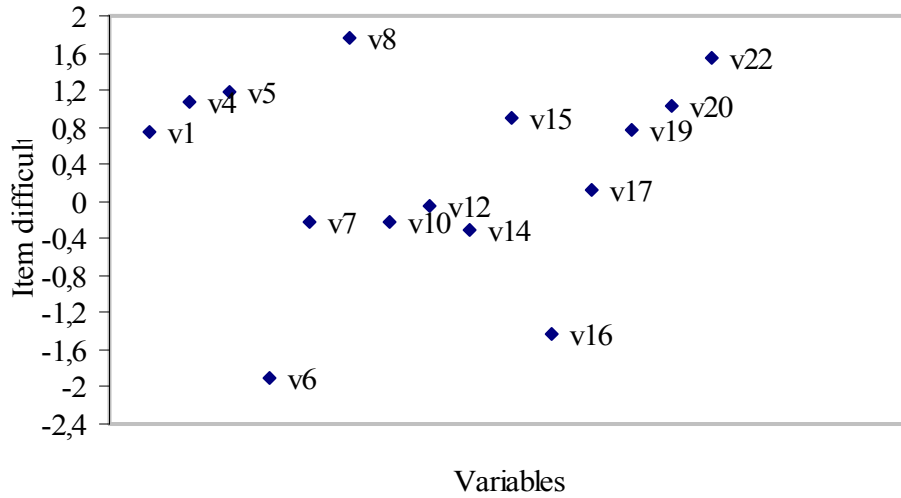
Figure 6.2 Variable-based item difficulty (VPA)
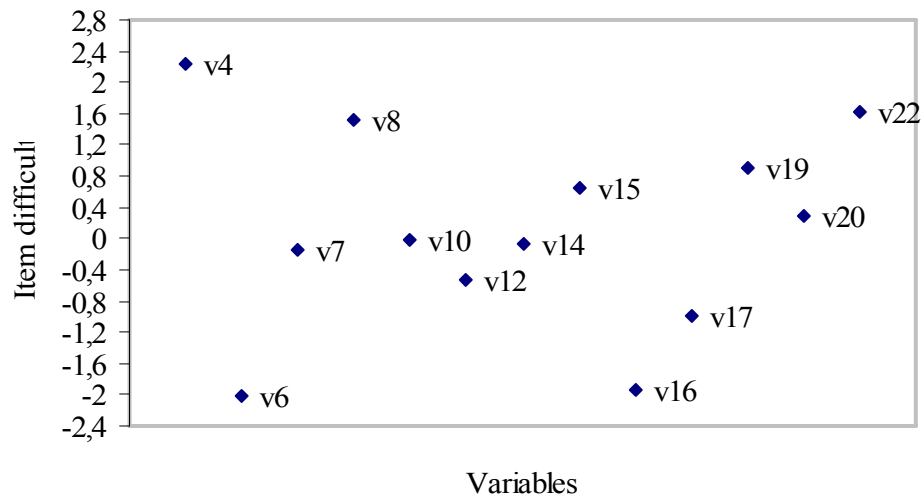
As shown in Figures 6.1 and 6.2, the hierarchical relationship among the variables in terms of the difficulty level reflects the impact of each variable as suggested by the results of our analysis in the previous section of this chapter (see Tables 6.9 and 6.10 Variable frequencies in Top vs Bottom groups of items). For instance, of the three item-type related variables, **v6** is at the bottom, **v8** is at or near the top of the hierarchy, while **v7** lies in between the two. This indicates considerable differences in the difficulty level of the three item types involved in the investigation. The position in the above diagrams of the three variables related to lexical overlap (**v16**, **v17** and **v19**) shows similar

differences in the difficulty level of items involving one or another of those variables. Linguistic characteristics of text / correct answer (**v1**), lower-frequency vocabulary in text / correct answer (**v4**), abstract topic (**v5**), as well as the two variables associated with the process of eliminating incorrect options (**v20** and **v22**) are all in the upper part of the hierarchy, which indicates their likely contribution to the difficulty of answering the items examined. When **v4** is compared to **v10**, we can see that, in both sets of data, v4 is much higher up in the hierarchy than v10.

However, when interpreting the results discussed above, it should be taken into account that an investigation into effects of item characteristics that is based on item statistics, as was the case with our analysis of variable-based item difficulty in this section, is not capable of accounting for the complex interactions among various item characteristics involved in responding to each individual item of the kind by each individual test taker, which we have earlier seen examples of in our Study Two on VPA. It is important to consider that, as Brindley and Slatyer (2002) point out, particular combinations of item characteristics may either accentuate or attenuate the effect on difficulty. 'Easy' or 'difficult' features may well cancel each other out (p.387). Besides, we should bear in mind that the actual difficulty of any test item is dependent of not only the characteristics of the items but also the characteristics of individual test takers, as we have seen in Study Two, and as will be shown from a different perspective in the section that follows, focusing on students' perception of the difficulty of the items.

6.3.2.3 Students' perception of item difficulty

In the third step of our examination of the relationship between item characteristic variables and item difficulty, the data obtained from student questionnaires was analysed. As mentioned in Study Two, at the end of each think-aloud session, students

were asked to assess the difficulty of each task and each item they had responded to on

a 1 – 6 point scale of difficulty, with 1 indicating the easiest, and 6 the most difficult

item according to students' perception. The ratings from the two students completing

the same items were analysed in relation to each other, the accuracy of answers, the

variables identified in each student's verbal protocol, and the IRT estimates of item

difficulty. Table 6.13 displays the data that formed the basis of the analysis.

Table 6.13 Students' perception of the difficulty of the items

Item	Observed variables	AC	R	Observed variables	AC	R	M
	Student 1 (LS)			**Student 2 (LMS)**			
1	v6, v10, v17	1	4	v6, v10, v21	1	5	-.69
2	v6, v16	1	1	v6, v16	1	1	-2.79
3	v6, v16	1	1	v6, v16	1	1	-2.41
4	v6, v16	1	2	v6, v16	1	4	-1.70
5	v10	9	6	v6, v16	1	1	-3.26
6	v6	1	1	v6, v16	1	1	-2.41
7	v6, v16	1	2	v6, v16	1	2	-2.69
8	v10	9	6	v10	0	5	-.80
9	v6, v10, v16	1	2	v6, v10, v16	1	2	-.93
10	v6, v16 , v17	1	5	-	0	1	-1.32
	Student 1 (MS)			**Student 2 (HS1)**			
11	v7, v14, v16	1	1	v7, v14, v20	1	2	-2.63
12	v7, v14, v16	1	1	v7, v15, v20	1	3	-1.42
13	v10	0	2	v7, v14	1	3	-.52
14	v7, v14, v16	1	1	v7, v14, v16	1	1	-1.98
15	v7	0	1	v7, v14	1	2	-.56
	Student 1 (LS)			**Student 2 (LMS)**			
16	v7	0	3	v7, v14	1	2	-1.28
17	v7, v10, v12	0	4	v7, v19	1	3	-1.23
18	v10, v12	9	6	v10, v12	1	3	-1.12
19	v10, v12	0	6	v10	9	6	-.51
20	v10, v12	9	6	v16	1	5	-.98
21	v10, v12, v16	1	5	-	0	3	-1.28
	Student 1 (MS)			**Student 2 (HS1)**			
22	v10, v12	0	5	v7, v15, v20	1	2	.44
23	v7	0	3	v7, v15, v20	1	2	-.16
24	v12	0	3	v7, v15, v20	1	2	-.12
25	v7	0	6	v7, v12, v15, v21	1	3	.51
26	v7, v15	1	5	v7, v15, v21	1	3	-.30
27	v7, v15, v16	1	5	v7, v15	1	1	-1.31
28	v15	1	4	v7, v15	1	1	-.37
	Student 1 (LMS)			**Student 2 (HS1)**			
29	v9	0	5	v9, v19	0	3	3.06
30	v9	9	6	v7, v14	0	6	2.26
31	-	9	6	v7, v10	0	6	.26
32	v7, v14	1	5	v9	0	6	1.36

33	v10	0	**4**	v7, v10, v14,	1	**2**	1.44
34	v10	9	**6**	v7, v14	1	**2**	.86
	Student 1 (MHS)			**Student 2 (HS2)**			
35	v10, v15	1	**3**	v8, v15, v20, v22	1	**3**	1.88
36	v8, v15, v20	1	**1**	v8, v15, v20, v22	1	**3**	1.96
37	v4, v9, v10, v15	0	**5**	v4, v10	0	**4**	2.53
38	v8, v15	1	**4**	v8, v15, v22	1	**3**	.89
39	v15	0	**2**	v8, v10, v15, v22	1	**4**	1.38
40	v4, v10, v15	1	**4**	v20, v22	0	**5**	1.98
41	v7, v15, v20	1	**1**	-	0	**3**	1.93
42	v7, v14	1	**1**	v14	1	**2**	1.66

Codes: AC=Answer Correct; 1=correct answer; 0=wrong answer; 9=blank;
R=Ratings given by students; M=Measure logit

When examining students' judgements of the difficulty of the items, Table 6.13 reveals that in the case of the vast majority of the items (76%), there was a difference between students' ratings on the same items. It would be reasonable to expect that, of the two students completing the same items, the lower-level student always perceived the items to be more difficult and, accordingly, assigned them higher ratings than the higher-level student. However, a closer examination of the table shows that to have happened, in a number of cases, the other way round. That is, the same item was judged to be easier by the lower-level student than the higher-level student and vice versa. For instance, **Item 4**, which was rated 2 (that is, easy rather than difficult) by the lower-level student (LS), was rated 4 (difficult rather than easy) by the higher-level student (LMS). **Items 11, 12, 13**, and **15** were all perceived to be more difficult by the higher-level student (HS1) than the lower-level student (MS). **Items 36, 39**, and **41** were again all judged to be easier by the lower-level student (MHS), whose ratings on these items included two 1s and a 2, than by the higher-level student (HS2), who gave the same items two 3s and a 4.

Looking at the ratings in relation to the accuracy of answers, we can see that, again contrary to expectations, in many cases, students did not perceive the items they had failed to answer to be more difficult than those they had answered correctly. On the contrary, they often judged an item they had failed to be easier than items they had been

able to answer correctly. For instance, **LMS**, despite her failure on the item, judged **Item 10** to be 'very easy', assigning it a rating of 1, while the same student judged **Item 1**, which she had been able to answer correctly, as 'rather difficult', assigning that item a 5. **MS**, despite her incorrect answers to **Items 23** and **24**, perceived both items to be easier than two other items on the same task, specifically, **Items 26** and **27**, both of which she had answered correctly. As Table 6.13 shows, she gave each of the former two items a 3, while each of the latter two a 5. **MHS** assigned **Item 39**, which she had failed, a 2 (judging the item easy rather than difficult), while both **Items 38** and **40** on the same task, which she had answered correctly, a 4 each (perceiving the latter two items to be difficult rather than easy).

When the ratings are examined with a view on individual variables identified in each student's verbal protocol, we can see that, apart from students' language level and/or the accuracy of their answers, the presence or absence of certain variables in the verbal reports offered, in many cases, reasonable accounts for the differences as well as for the agreement between ratings of either the same student or of the two students completing the same items.

For instance, in the case of **Item 8**, Table 6.13 shows that the lower-level student (LS) failed to give any answer to the item, while the higher-level student (LMS) gave a wrong answer. Both students judged the item (very/rather) difficult, with LS assigning it a 6, while LMS a 5. The only variable identified in their verbal reports is v10, which indicates that both students had difficulties in understanding key vocabulary in the question and/or the correct answer. Looking back to Study Two, the analysis of the verbal reports reveals that, when completing the item, both students lacked knowledge of the key word 'entertainment' used in the question. However, the higher-level student

(LMS) took the word to mean 'furniture' and, accordingly, was able to select from among the options an answer that she had thought would fit in with the meaning of that word. Although, as the analysis also reveals, she selected her answer with not much confidence, the fact that, unlike LS, she was able to give some answer (even if a wrong one), might have motivated her to assign the same item a slightly lower rating than the other student.

Item 1 provides a different example of the relationship between the variables and students' perceptions of the difficulty of the items. As we can see from Table 6.13, this item was answered correctly by both students, yet was rated 4 by LS, and 5 by LMS. Apart from the item-type related variable (v6), the only variable common to both students' verbal reports regarding this item is v10, indicating vocabulary problems on the part of both students. While v10 might explain the relatively high difficulty ratings on this item, the occurrence of a third variable in the verbal report of each student may account for the difference between students' ratings. In LS' verbal report, the third variable involved is v17, while in LMS' protocol it is v21. If we look at our earlier **Figures 6.1 and 6.2**, which display variable-based item difficulties, we can see that the average difficulty level of items associated with v17 is lower, in both the Content Analysis and VPA data, than that of v20 (which, it should be noted, resulted from merging v20 and v21). In other words, LMS appears to have arrived at the correct answer in a more difficult way than LS, which might be a possible reason why she judged this item to be slightly more difficult than LS did.

Similar reasons may explain the difference between the two students' ratings on the difficulty of **Items 11 and 12**, both of which were answered correctly by both students, yet were rated to be more difficult by the higher-level student (HS1) than by the lower-

level student (MS). Examining the variables occurring in this item in each student's verbal protocol, it can be seen that MS chose an easier way to answer the item, relying on the lexical overlap between the item and the correct answer (v16) in the case of both items, which may explain why she judged both items 'very easy', assigning each a rating of 1. In contrast, the occurrence of v20 in the other student's (HS1) verbal report suggests that HS1 selected her answer to the same items by eliminating incorrect options through comparing their meanings and making inferences on the basis of information in those options, which is apparently a more tedious way to answer the items in question than using a word-matching strategy. With this considered, her higher ratings of a 2 and a 3 on the same items, despite her higher language proficiency, appear to be reasonable.

From a different perspective, from Table 6.13 it can also be seen that students' ratings on the items differed from the IRT estimates of difficulty in some cases to a smaller, in others, to a greater extent. Looking at the easiest and most difficult items among those involved in the investigation, we can see that the easiest item (Item 5), with a logit value of -3.26, was rated 1 by one of the two students, which reflects an agreement with the difficulty estimate for the item, unlike the rating of 6 given by the other student on the same item. The most difficult item (Item 29), with a logit score of 3.06, received a rating of 3 from one student, and 5 from the other.

In order to obtain a measure of the strength of association between ratings of the two students completing the same items, on the one hand, and ratings of each student and the item difficulty estimates, on the other, correlations were examined by task. The results of this analysis are summarised in Table 6.14.

Table 6.14 Correlations between students' ratings and difficulty estimates, by task

Task 1

	M	*RS2*
RS1	0.29	0.34
RS2	0.70	

Task 2

	M	*RS2*
RS1	0.55	0.53
RS2	0.47	

Task 3

	M	*RS2*
RS1	0.63	0.73
RS2	0.92	

Task 4

	M	*RS2*
RS1	0.16	0.36
RS2	0.63	

Task 5

	M	*RS2*
RS1	-0.28	0.44
RS2	-0.14	

Task 6

	M	*RS2*
RS1	0.14	0.59
RS2	0.28	

Codes: M=Measure logits (IRT estimates of item difficulty); RS1=Ratings from Student 1;
RS2=Ratings from Student 2;

Concerning the relationship between *ratings from the two students on the same tasks*, the rather low correlation coefficients in the case of Task 1 (r=0.34), Task 4 (r=0.36), and Task 5 (r=0.44) indicate that the two students completing the same task perceived the difficulty of the items in the tasks in question very differently from each other. The ratings converged to the greatest degree in the case of Task 3, where the correlation coefficient can be said to be quite high (r=0.73). With respect to the *relationship of* each student's *ratings to the empirical estimates of item difficulty*, the results showed even

greater variations, ranging from the very strong agreement of r=0.92 in the case of Task 3 by Student 2, through the rather weak (occasionally close to complete lack of) agreement reflected in the correlations of r=0.16, r=0.14, r=0.29 and r=0.28 in the case of Task 1, Task 4, and Task 6, to negative correlations in the case of Task 5 (r=-0.28 and r=-0.14), with the latter figures indicating that items on the task shown by the Measure logits to be more difficult than the others were, in many cases, perceived to be easier by both students and vice versa, i.e., items shown by the difficulty estimates to be easier than others were perceived to be more difficult by the students.

Overall, the results of the analyses of students' perception of item difficulty suggest that, on the one hand, the difficulty of the items was, more often than not, perceived very differently by the two students responding to the same items and, on the other, there was very weak agreement between item difficulties as perceived by the students participating in the study and as measured by the IRT estimates of item difficulty.

6.4 Summary and conclusion

In this chapter, aiming to find answers to our Research Questions 3 and 4, we explored the relationship between findings of Study One and Study Two, on the one hand, and the item characteristic variables identified and the difficulty of the items involved in the investigation, on the other. The issue of the agreement between Content Analysis and VPA (RQ3) was examined in two different ways. One was based on a comparison of data from the two studies, with a focus on evidence in the verbal protocols of the main process, skill, or difficulty predicted by Content Analysis to be involved in each item, while the second approach involved an examination of the same issue from the perspective of the 22 item characteristic variables identified in Study One.

Results of the comparison carried out in the first case indicated a 56% agreement between findings from the two studies, which was confirmed by the results of the analysis focusing on the agreement between predicted and observed frequencies of the occurrence of each variable in the set of items examined. From the analysis of the relationship between predicted and observed variable occurrences it has become clear that certain item characteristics (mainly but not exclusively text-related features), however crucial part they might play in the difficulty of answering the items, are unlikely to occur in verbal protocol data. With such variables excluded from the analysis, the agreement between predicted and observed frequencies of the variables increased to 77%, which implies that the majority of the skills and processes that the subjects participating in the study used in actual completion of the items had been successfully specified employing the methodology of Content Analysis.

Our examination of the match between predicted and observed frequencies of each variable resulted in merging and/or discarding some of the original 22 variables for the purpose of the next stage of the investigation, which focused on issues of the relationship between the item characteristic variables and the difficulty of the items, that is, issues raised by our Research Question 4. Of the reduced number of 15 variables, nine appeared to have notable impact on the difficulty of the reading items examined. Seven of them were observed to make the items more difficult, while two variables were evidenced to have 'easifying' effects. An examination of variable-based item difficulties revealed that, according to both Content Analysis and VPA data, there was a hierarchical relationship among the variables in terms of the difficulty level.

However, the results of the investigation in this study should be interpreted in their rightful context. Apart from limitations related to methodological issues that were

referred to at various points of the analysis, our investigation into students' perceptions of item difficulty showed that the actual difficulty of any test item depends on the characteristics of the test taker, and not just on the characteristics of the item. Therefore, the results emerging from this study should be seen as reflecting tendencies rather than solid claims regarding effects of the item characteristic variables identified on the difficulty of such reading test items in general.

Chapter 7 Discussion and conclusions

This dissertation examined and explored effects of task and item features on learners' performance on EFL reading comprehension tests, with a focus on characteristics of matching tasks. The research was motivated by the fact that, although such tasks are commonly used in recent tests of second and foreign language reading comprehension, including the new Hungarian school-leaving examination, the effects of item characteristic variables specific to this particular type of reading tasks have not been in the focus of attention in any previous studies investigating factors underlying performance on reading tests. The main purpose of the research was to identify item characteristics likely to influence learners' scores on such tasks and items, and examine the effects of the variables identified on the difficulty of the tasks and items involved in the investigation.

It was felt useful to use a triangulation approach to exploring the issue as it enabled an examination of the relationships among different sources of information, specifically, information obtained from content analyses of the tasks and items, verbal report data on the cognitive processes, skills, and knowledge involved in actual completion of the tasks, item statistics on the difficulty of the items, and student questionnaires on perceived item difficulties. Relating different types of information to one another, and accumulating evidence from various sources, helps provide a better understanding of the interactions among the variables believed to affect the difficulty of such reading items, the processes test takers actually use in carrying out such tasks, and learners' performance on such tasks and items.

The results of the three studies suggest that the reading items examined share certain common features with the traditional 4-option multiple-choice questions investigated in previous studies, while, at the same time, some characteristics identified appear to be specific to the type of reading items focused on in this research. For example, variables associated with lexical overlap, which have been observed in previous studies (Freedle and Kostin 1993; Buck et al. 1997) to relate to the difficulty of answering multiple-choice items were observed to impact on the difficulty of these reading items as well. On the other hand, for instance, variables related to the elimination of incorrect options, which have received very little attention in previous research studies, were evidenced to play a particularly emphatic part in the difficulty of answering many of the items examined here. The results of all three studies carried out in this research suggest that, in the case of the reading items examined, because of the high number of (7 to 10, and in the case of one task, 16) options from which test takers have to choose their answers, the complex process of eliminating incorrect options will significantly increase the demands of answering the items. The most important results from each study are summarized in the section that follows, before discussing limitations of the investigation and implications for further research.

7.1 Summary of the results

The purpose of Study One was to identify variables likely to affect the difficulty of the reading items under investigation. For this purpose, a detailed description of the content of the tasks and items was carried out, using a modified version of Bachman and Palmer's (1996) framework of language task characteristics. Based, in part, on the information obtained from the analysis of the tasks and items and, in part, on theoretical models of reading, and the research literature relevant to the issue, an initial list of 36 item characteristic variables believed to impact on the difficulty of the items was drawn

up, which was then modified and revised several times before the final set of 22 variables was used to code each item on each task for the variables involved in their completion. The results of coding showed a rather uneven distribution of the variables across the tasks and items in terms of both the frequency of their occurrences and the particular combinations in which they occurred in individual items. While certain combinations of variables appeared to characterize most items in one particular task, items in another task were observed to involve other combinations of variables, reflecting considerable variation in the construct measured by individual items across the six tasks examined.

Study Two employed verbal protocol analysis to explore the variables involved in responding to the items from the perspective of the test taker. The analysis of verbal report data revealed similarities as well as great differences in both students' overall approach to processing the texts and tasks, and the skills, processes and strategies they had used in producing their answers to individual items. Of the six students participating in the study, one lower-level student generally read and attempted to translate the text word by word, while the other students typically processed the text section by section, tried to understand it in only as much detail as they thought was necessary to answer the items, and they generally paraphrased or summarized what they had understood. In terms of task processing, it was observed that, contrary to expectations encouraged by theoretical definitions/descriptions of processing reading tasks, students generally did not read through the whole text to get an overall picture of what the text was about before embarking on answering the items on the text. There were considerable differences in the approaches and strategies students used in responding to the items. The most striking differences observed relate to

- how much time students spent trying to select their answer to a particular item before going on to read the next section of the text which involved the next item on the task;

- whether or not they read either sections of the text or the options carefully when a correct answer required careful reading for details;

- how systematic they were in checking the suitability of their selected or intended answers;

- whether they were able to guess the meaning of unfamiliar words and phrases crucial for a correct answer in either the text or the options, including the correct option;

- whether they were able to eliminate (an) incorrect option(s) that had a semantic overlap with either the correct answer or the relevant section of the text;

- the extent to which they relied on the lexical overlap between the item and the correct answer or incorrect options when selecting their answers.

The qualitative analysis of verbal reports provided rich descriptive data on just what particular combinations of item characteristics made certain items easier or more difficult to answer for the students participating in the study. It has revealed that students, despite demonstrating the skill or knowledge required by an item, not infrequently failed to select the correct answer to the item, while there were cases when they answered the item correctly despite an apparent failure to understand the meaning of relevant sections of the text. There was ample evidence in the verbal protocols for students' arriving at the same correct or incorrect answer in very different ways. Overall, the verbal report data provided very useful insights into the processes test takers used when completing the items.

Study Three brought together all available information on the item characteristic variables underlying responses to the items, on the one hand, and the difficulty of the items, on the other. The main purpose of Study Three was to find out, first, whether and

to what extent the data analyses in Study One and Study Two revealed similar findings and, second, whether there was a relationship between the item characteristics identified and the difficulty of the items. With respect to the first issue, the results of the analyses showed a 56% agreement between findings from the two studies. When variables that are less likely to occur in verbal report data (e.g., the length and syntactic complexity of sentences) were excluded from the analysis, the agreement between predicted and observed frequencies of variable occurrence increased to 77%, which can be considered to be a relatively high prediction rate.

In light of the results of the analysis of the relationship between findings from the content analysis and VPA studies, it was found useful to merge or discard some of the original 22 item characteristic variables for the purpose of exploring the effects of a reduced number of variables on the difficulty of the items. Eventually, 15 variables resulting from merging were included in the investigation into the relationship between the variables identified and item difficulty, seven of which were found to make the items more difficult, while two variables proved to make the items easier to answer. This finding was confirmed by the results of an examination of variable-based item difficulties, which showed a hierarchical relationship among the variables in terms of the difficulty level. The average difficulty level of items associated with particular variables was notably higher or lower than that of the items not involving those variables.

Study Three also investigated students' perceptions of the difficulty of the items. The correlations between ratings from the two students completing the same tasks, calculated for the whole item set, indicated a rather weak agreement (r=0.43) between students' perceptions of the difficulty of the same items. The correlations regarding the

relationship between students' ratings of the difficulty of the items and the IRT estimates of difficulty showed a similarly weak agreement, indicating considerable differences in item difficulties as measured by difficulty estimates and as perceived by the students participating in the study.

In sum, nine of the final set of 15 item characteristic variables that were identified were observed to have notable effects on the difficulty of the reading items examined. The investigation showed that many of the variables underlying performance on the items could be identified using the methodology of content analysis. On the other hand, the use of a triangulation approach to exploring the issue, in particular, the verbal report data in Study Two and the data on students' perception of item difficulty in Study Three, made it clear that to be able to determine the actual difficulty of the items we would need to be able to describe the possible interactions among the item characteristic variables for each individual test taker. Bachman (2000) remarked on the issue as follows.

> As soon as one considers what makes items difficult, one immediately realizes that difficulty isn't a reasonable question at all. A given task or item is differentially difficult for different test takers and a given test taker will find different tasks differentially difficult. *Ergo*, difficulty is not a separate quality at all, but rather a function of the interaction between task characteristics and test taker characteristics. When we design a test, we can specify the task characteristics, and describe the characteristics of the test takers, but getting at the interaction is the rub.
> (Bachman 2000, cited in Brindley and Slatyer 2002: 390)

All this suggests that, for one thing, the results of this investigation should be treated as providing, at best, some preliminary evidence for relationships between the variables identified and the difficulty of such reading items and, for another, continued research relying on introspective and content analysis data is required to further explore the effects of item characteristic variables on the difficulty of reading test items, in general, and the type of reading items investigated here, in particular.

7.2 Limitations of the research

There are several limitations to consider when interpreting the results of this investigation. First, it is not advisable to generalize the results to either reading tasks or student population different from those involved in this research. Neither the tasks nor the subjects providing verbal report data were selected to provide representative samples. The same tasks completed by other students with different language proficiency levels and other test taker characteristics, as well as the same subjects completing other tasks of the same type investigated here might yield results, to a smaller or greater degree, different from those reported on here. Second, the generalizability of the results is also limited by the small sample of subjects involved in the study exploring verbal report data. Third, as mentioned earlier, the development of the framework of item characteristics in Study One was not without problems. One of the difficulties encountered in the development process was in operationalizing some of the item characteristics identified, and another difficulty resulted from the variable use in the literature of terminology and certain concepts relevant to the research. Fourth, regarding the procedure of coding the items, unfortunately, no resources were available to involve a second coder and check inter-coder reliability. Fifth, considering the use of verbal report data, apart from the limitations discussed in Study Two, it should be noted that it is very likely that there were processes used that were not reported by the subjects completing the tasks. It must be very difficult to verbalise one's thoughts while processing a reading test task. Lastly, it might also be worth considering that the skills and processes underlying students' correct as well as incorrect answers were coded and, likewise, no distinction was made between the variables in this respect in the analysis of the data, either. It is possible that an analysis of the verbal report data with the inclusion of only the variables underlying correct answers would have led to different results.

7.3 Implications for further research

Further research into task and item effects, with a similar interest in factors affecting the difficulty of matching items, could include, for one thing, other types of analyses conducted on the data collected for purposes of this research. For example, as hinted above, by examining only the variables involved in items that test takers answered correctly, it might be possible to obtain a clearer picture of what combinations of variables underlie successful completion of the items.

Furthermore, as in this research there were only six Hungarian secondary school students providing verbal report data on the items, our study could be replicated with the involvement of a larger sample of subjects, which could provide additional information on the variables involved in actual completion of these items. Preferably, replication studies would involve subjects with a range of different test taker characteristics, including different age groups, language levels, and nationalities/first language backgrounds.

In addition to exploring the data on the tasks and items examined here, it would be very useful to collect similar data on other matching tasks, possibly from multiple sources, and examine if the same variables reported on here affected learners' performance on other matching tasks.

Lastly, the framework of item characteristics developed and used to code the items in this research could be further refined and tried out on other tasks to see if the descriptions of the variables could be reliably used to code and examine factors underlying test performance on a range of other matching items.

7.4 Conclusion

This dissertation investigated item characteristic variables that affect learners' performance on matching reading test items developed by the Hungarian Examinations Reform Project for purposes of the new Hungarian school-leaving examination in English. The investigation relied on various sources of data to determine item characteristics that may account for differences in the difficulty of these reading items. In line with research studies suggesting the use of a triangulation approach to exploring the relationship between task and item features and learners' scores on reading comprehension test (e.g., Anderson et al. 1991; Gao and Rogers 2007), this thesis highlighted both the value of using and the need to use, whenever possible, multiple sources of data in the investigation of task and item difficulty. The research findings from the three studies reported on in this thesis are hoped to provide useful information for language testers developing matching reading test tasks, in particular, in the Hungarian context, but also in other EFL reading assessment contexts similar to the Hungarian one. From a different perspective, it is hoped that this dissertation, by examining and exploring task and item features that may impact on test takers' performance on this particular type of reading test tasks, will contribute to a better understanding of the nature and effects of factors that underlie performance on reading comprehension tests in general.

References

Ábrahám, K., and Jilly, V. (1999) The School-leaving Examination in Hungary. In Fekete, H., Major, É., and Nikolov, M. (1999) (eds.) *English Language Education in Hungary. A Baseline Study.* Budapest: The British Council Hungary. 21-53.

Alderson, J. C. (1984) Reading in a foreign language: a reading problem or a language problem? In Alderson, J. C. & Urquhart, A. H. (eds.) pp. 1-24.

Alderson, J. C. (1990a) Testing Reading Comprehension Skills (Part One). *Reading in a Foreign Language*, 6 (2), 425-438.

Alderson, J. C. (1990b) Testing Reading Comprehension Skills (Part Two). Getting Students to Talk About Taking a Reading Test. (A Pilot Study) *Reading in a Foreign Language*, 7 (1), 465-503.

Alderson, J. C. (2000) *Assessing Reading*. Cambridge: Cambridge University Press.

Alderson, J.C., and Banerjee, J. (2001) Language Testing and Assessment (Part 1). *LanguageTeaching*, 34, 213-236. Cambridge University Press.

Alderson, J.C., and Banerjee, J. (2002) Language Testing and Assessment (Part 2). *Language Teaching*, 35, 79-113. Cambridge University Press.

Alderson, J. C., Clapham, C., and Wall, D. (1995) *Language test construction and evaluation*. Cambridge: Cambridge University Press.

Alderson, J. C., and Lukmani, Y. (1989) Cognition and Reading: Cognitive Levels as Embodied in Test Questions. *Reading in a Foreign Language*, 5 (2), 253-270.

Alderson, J. C., and Urquhart, A. H. (eds.) (1984) *Reading in a Foreign Language*. London: Longman.

Alderson, J. C., and Urquhart, A. H. (1985) This test is unfair: I'm not an economist. In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.) (1988) pp. 168-182.

Alderson, J. C., Nagy, E., and Öveges, E. (2000) (eds.) *English Language Education in Hungary. Part II Examining Hungarian Learners' Achievements in English.* Budapest: The British Council Hungary.

Anderson, N. J., Bachman, L., Perkins, K., and Cohen, A. (1991) An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources. *Language Testing*, 8 (1), 41-66.

Bachman, L. F. (1990) *Fundamental considerations in language testing*. Oxford: Oxford University Press.

Bachman, L. F., Davidson, F., and Milanovic, M. (1996) The use of test method characteristics in the content analysis and design of EFL proficiency tests. *Language Testing*, 13 (2), 125-150.

Bachman, L. F., and Palmer, A. S. (1996) *Language testing in practice*. Oxford: Oxford University Press.

Bárány, F., Major, É., Martsa S., Martsa S. É., Nagy, I., Nemes, A., Szabó, T., and Vándor, J. (1999) Stakeholders' Attitudes. In Fekete, H., Major, É., and Nikolov, M. (eds.) (1999) pp. 137-204.

Bartlett, F. C. (1932) *Remembering*. Cambridge: Cambridge University Press.

Beaugrande, R. de. (1982) The story of grammars and the grammar of stories. *Journal of Pragmatics*, 6, 383-422.

Beaugrande, R. de., and Dressler, W. U. (1981) *Introduction to Text Linguistics*. London: Longman.

Bernhardt, E. B. (1991a) *Reading Development in Second Language: Theoretical, Empirical and Classroom Perspectives*. New Jersey: Ablex Publishing Corporation.

Bernhardt, E. B. (1991b) A psycholinguistic perspective on second language literacy. In Hulstijn, J.H., and Matter, J.F. (eds.) (1991) Reading in two languages. *AILA Review*, 8 (Amsterdam), 31-44.

Brindley, G., and Slatyer, H. (2002) Exploring task difficulty in ESL listening assessment. *Language Testing*, 19 (4), 369-394.

Brown, G., and Yule, G. (1983) *Discourse Analysis*. Cambridge: Cambridge University Press.

Buck, G. (1991) The testing of listening comprehension: an introspective study. *Language Testing*, 8 (1), 67-91.

Buck, G. (1994) The appropriacy of psychometric measurement models for testing second language listening comprehension. *Language Testing*, 11, 145-170.

Buck, G. (2001) *Assessing Listening*. Cambridge: Cambridge University Press.

Buck, G., and Tatsuoka, K. (1998) Application of the rule-space procedure to language testing: examining attributes of a free response listening test. *Language Testing*, 15 (2), 119-157.

Buck, G., Tatsuoka, K., and Kostin, I. (1997) The Subskills of Reading: Rule-space Analysis of a Multiple-choice Test of Second Language Reading Comprehension. *Language Learning*, 47:3, 423-466.

Canale, M. (1983a) From communicative competence to communicative language pedagogy. In Richard, J. C., and Schmidt, R. W. (eds.) (1983) *Language and Communication*. London: Longman. pp. 2-27.

Canale, M. (1983b) On some dimensions of language proficiency. In Oller, J. W. (ed.) *Issues in Language Testing Research*. Newbury House, Rowley, MA. pp. 333-342.

Canale, M., and Swain, M. (1980) 'Theoretical bases of communicative approaches to second language teaching and testing.' Applied Linguistics, 1 (1), 1-47.

Carpenter, P. A., and Just, M. A. (1975) Sentence comprehension: a psycholinguistic processing model of verification. *Psychological Review*, 82, 45-73.

Carrell, P. L. (1988) Introduction: Interactive approaches to second language reading. In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.) pp. 1-7.

Carrell, P. L. (1988) Some causes of text-boundedness and schema interference in ESL reading. In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.) pp. 101-113.

Carrell, P. L., and Eisterhold, J. C. (1983) Schema theory and ESL reading pedagogy. *TESOL Quarterly*, 17, 553-573.

Carrell, P. L., Devine, J., and Eskey, D. E. (1988) (eds.) *Interactive Approaches to Second Language Reading*. Cambridge: Cambridge University Press.

Carver, R. P (1982) Optimal rate of reading prose. *Reading Research Quarterly*, XVIII (1), 56-88.

Carver, R. P. (1983) Is reading rate constant or flexible? *Reading Research Quarterly*, XVIII (2), 190-215.

Carver, R. P. (1984) Rauding theory predictions of amount comprehended under different purposes and speed reading conditions. *Reading Research Quarterly*, XIX (2), 205-218.

Chapelle, C. A. (1999) Validity in Language Assessment. *Annual Review of Applied Linguistics*, 19, 254-272. Cambridge: Cambridge University Press.

Chomsky, N. (1965) *Aspects of the theory of syntax*. Cambridge, Mass.: MIT Press.

Clapham, C. M. (1996) *The development of IELTS : a study of the effect of background knowledge on reading comprehension*. Cambridge: Cambridge University Press.

Clarke, M. A. (1988) The short circuit hypothesis of ESL reading – or when language competence interferes with reading performance. In Carrell, P. L., Devine, J., Eskey, D. E. (eds.) pp. 114-124.

Coady, J. (1979) A psycholinguistic model of the ESL reader. In R. Mackay, B. Barkman, and R.R. Jordan (eds.) *Reading in a second language*, 5-12. Roley, Mass.: Newbury House.

Cohen, A. (1998) *Strategies in learning and using a second language*. London: Longman.

Cohen, A., and Upton, T. (2006) *Strategies in Responding to the New TOEFL Reading Tasks*. Monograph Series. ETS. MS-33, April 2006. RR-06-06.

Colby, B. (1982) Notes on the transmission and evolution of stories. *Journal of Pragmatics*, 6, 463-472.

Council of Europe (2001) *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Strasbourg: Council for Cultural Co-operation, Education Committee. Cambridge: Cambridge University Press.

van Dijk, T., and Kintsch, W. (1983) *Strategies of discourse comprehension*. New York: Academic Press.

Dörnyei, Z., Nyilasi, E., and Clément, R. (1996) Hungarian school children's motivation to learn foreign languages: A comparison of target languages. *NovELTy*, 3 (2), 6-16.

Dörnyei, Z., and Schmidt, R. (eds.) (2000) *Motivation and second language acquisition*. Honolulu. HI: The University of Hawaii. Second Language Teaching and Curriculum Center.

Ericsson, K. A., and Simon, H. (1993) *Protocol Analysis*. Cambridge. Mass: MIT Press.

Eskey, D. E. (1988) Holding in the bottom: an interactive approach to the language problems of second language readers. In Carrell, P. L., Devine, J., and D. E. Eskey (eds.) (1988).

Eskey, D., and Grabe, W. (1988) Interactive models for second language reading: perspectives on instruction. In Carrell, P. L., Devine, J., and D. E. Eskey (eds.).

Fekete, H., Major, É., and Nikolov, M. (1999) (eds.) *English Language Education in Hungary. A Baseline Study*. Budapest: The British Council Hungary.

Freedle, R., and Kostin, I. (1991) *The prediction of SAT reading comprehension item difficulty for expository prose passages*. Princeton, NJ: ETS Research Report RR-91-29.

Freedle, R., and Kostin, I. (1992) *The prediction of GRE reading comprehension item difficulty for expository prose passages for each of three item types: main ideas, inferences and explicit statements*. Princeton, NJ: ETS Research Report RR-91-59.

Freedle, R., and Kostin, I. (1993) The prediction of TOEFL reading item difficulty: implications for construct validity. *Language Testing,* 10, 133-170.

Fries, C. C. (1963) Linguistics and Reading. New York: Holt, Rinehart and Winston.

Gao, L., and Rogers, T. (2007) *Cognitive-Psychometric Modeling of the MELAB Reading Items*. University of Alberta. A paper prepared for presentation at the annual meeting of the National Council of Measurement in Education, Chicago, Illinois, April 2007.

Gass, S. M., and Mackey, A. (2000) *Stimulated Recall Methodology in Second Language Research*. Lawrence Erlbaum Associates.

Goodman, K. S. (1967) Reading: a psycholinguistic guessing game. *Journal of the Reading Specialist*, 6 (1), 126-135.

Goodman, K. S. (1971) Psycholinguistic universals in the reading process. In P. Pimsleur and T. Quinn (eds.) *The psychology of second language learning*, 135-142. Cambridge: Cambridge University Press.

Goodman, K. S. (1988) The reading process. In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.) (1988) pp.11-21.

Gough, P. B. (1972) One second of reading. In Kavanagh, J. F. and I. G. Mattingley (eds.) *Language by Ear and Eye*. Cambridge, Mass.: MIT Press.

Grabe, W. (1988) Reassessing the term "interactive". In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.). pp 56-70.

Grabe, W. (1991) Current developments in second-language reading research. *TESOL Quarterly*, 25 (3), 375-406.

Grabe, W. (2000) Developments in reading research and their implications for computer-adaptive reading assessment. In M. Chalhoub-Deville (ed.) *Issues in computer-adaptive tests of reading*. Cambridge: Cambridge University Press.

Green, A. (1998) *Verbal protocol analysis in language testing research: A handbook*. Cambridge: University of Cambridge Local Examinations Syndicate.

Grotjahn, R. (1986) 'Test validation and cognitive psychology: some methodological considerations.' *Language Testing*, 3, 2, 159-185.

Hae-Jin, K., Jasuyo, S., and C. Gentile (2007) *Q-matrix construction: Defining the link between constructs and test items in cognitive diagnostic approaches*. Paper presented at the Language Testing Research Colloquium (LTRC), June 9-12, 2007. Barcelona, Spain.

Halliday, M. A. K. (1970) Language structure and language function. In J. Lyons (ed.) *New Horizons in Linguistics*, 140-165. Harmondsworth: Penguin.

Halliday, M. A. K. (1973) 'Towards a sociological semantics'. In Brumfit, C. J. and Johnson, K. (eds.) (1979) *The communicative Approach to Language Teaching*, 27-45. Oxford: Oxford University Press.

Halliday, M. A. K. (1975) *Learning How to Mean: Explorations in the Development of Language*. London: Edward Arnold.

Halliday, M. A. K. (1989) *Spoken and Written Language*. Oxford: Oxford University Press.

Halliday, M. A. K., and R. Hasan (1976) *Cohesion in English*. London: Longman.

Hare, V., Rabinowitz, M. and Schieble, K. (1989) Text effects on main idea comprehension. *Reading Research Quarterly*, 24, 72-88.

Hatch, E. (1992) *Discourse and Language Education*. Cambridge: Cambridge University Press.

Hatch, E., and Lazaraton, A. (1991) *Design and Statistics for Applied Linguistics. The Research Manual*. Boston: Heinle & Heinle Publishers.

Hoey, M. (1983) *On the Surface of Discourse*. London: Allen and Unwin.

Hoey, M. (1991) *Patterns of Lexis in Text*. Oxford: Oxford University Press.

Hosenfeld, C. (1984) Case studies of ninth grade readers. In Alderson, J. C. and Urquhart, A. H. (eds.) *Reading in a Foreign Language*, 231-249. London: Longman.

Hymes, D. (1972a) 'On Communicative Competence.' In J. Pride and Holmes, J (eds.) (1972) *Sociolinguistics: selected readings*. Harmondsworth: Penguin: 269-293.

Hymes, D. (1972b) Models of the interaction of language and social life. In Gumperz, J., and Hymes, D. (eds.) *Directions in Sociolinguistics: the Ethnography of Communication*, 35-71. New York: Holt, Rinehart and Winston.

Hymes, D. (1974) Toward ethnographies of communication. In *Foundations in Sociolinguistics: an Ethnographic Approach*, 3-28. Philadelphia: University of Pennsylvania Press.

Jang, E. E. (2005) *A validity narrative: Effects of reading skills diagnosis on teaching and learning in the context of NG TOEFL*. Doctoral dissertation. University of Illinois at Urbana Champaign. Urbana, Illinois.

Johnson, P. (1982) Effects on reading comprehension of building background knowledge. *TESOL Quarterly*, 16 (4), 503-516.

Kádárné, F. J. (1979) Az angol nyelv tanításának eredményei. In Kiss, A., Nagy, S., and Szarka, J. (eds.) *Tanulmányok a neveléstudomány köréből 1975-76*. Budapest: Akadémia. pp. 276-341.

Kieras, D. E. (1985) Thematic processes in the comprehension of technical prose. In Britton, B. and Black, J., editors, *Understanding expository text*. Hillsdale, NJ: Lawrence Erlbaum.

Kintsch, W., and van Dijk, T. A. (1978) Toward a Model of Text Comprehension and Production. *Psychological Review*. Vol 85 Number 5, 363-394.

LaBerge, D., and Samuels, S. J. (1974) Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 6, 293-323.

Linde, C., and Labov, W. (1975) Spatial networks as a site for the study of language and thought. *Language*, 51, 924-939.

Mandler, J. M. (1978) A code in the node: the use of a story schema in retrieval. *Discourse Processes*, 1, 14-35.

Mandler, J. M., and Johnson, N. S. (1977) Remembrance of things parsed: story structure and recall. *Cognitive Psychology*, 9, 111-151.

McCarthy, M. (1991) *Discourse Analysis for Language Teachers*. Cambridge: Cambridge University Press.

McCarthy, M, and Carter, R. (1994) *Language as Discourse: Perspectives for Language Teaching*. London: Longman.

Meehan, J. R. (1982) Stories and cognition: comments on Robert de Beaugrande's 'The story of grammars and the grammar of stories'. *Journal of Pragmatics*, 6, 455-462.

Messick, S. (1995) Validity of Psychological Assessment. Validation of Inferences From Persons' Responses and Performances as Scientific Inquiry Into Score Meaning. *American Psychologist*, 50, 9, 741-749.

Messick, S. (1996) Validity and washback in language testing. *Language Testing*, 13 (3), 241-256.

Meyer, B., and Freedle, R. (1984) The effects of different discourse types on recall. *American Educational Research Journal*, 21, 121-143.

Minsky, M. (1975) A framework for representing knowledge. In *The psychology of computer vision*, P.H. Winston (ed.), 211-277. New York: McGraw-Hill.

Munby, J. (1978) *Communicative syllabus design*. Cambridge: Cambridge University Press.

Nagy, E. (2000) A Chronological Account of the English Examination Reform Project: the Project Manager's Perspective. In Alderson, J. C., Nagy, E., and Öveges, E. (eds.) (2000) pp. 22-37.

Nash, W. (1985) *The language of humour. Style and technique in comic discourse*. London and New York: Longman.

Nikolov, M. (1999a) The Socio-Educational and Sociolinguistic Context of the Examination Reform. In Fekete, H., Major, É., and Nikolov, M. (eds.) (1999) pp. 7-20.

Nikolov, M. (1999b) "Why do you learn English?" "Because the teacher is short." A study of Hungarian children's foreign language learning motivation. *Language Teaching Research*, 3 (1), 33-56.

Nikolov, M. (2001a) Test-taking strategies of 12-year-old Hungarian learners of English as a foreign language. Paper presented at EARLI Conference, Fribourg, Switzerland.

Nikolov, M. (2001b) Hatodikosok feladatmegoldó stratégiái olvasott szöveg értését és íráskészséget mérő feladatokon angol nyelvből. Paper presented at Neveléstudományi Konferencia, MTA, Budapest, 2001. október.

Nikolov, M. (2001c) Minőségi nyelvoktatás – a nyelvek európai évében. *Iskolakultúra*, 8, 3-12.

Noijons, J., and Nagy, E. (1995) *Towards a standardised examinations system. Joint Hungarian-Dutch Project*. Final reports. Budapest: CITO, OKI.

Pearson, P. D., and Johnson, D. D. (1978) *Teaching reading comprehension*. New York: Holt, Rinehart and Winston.

Polanyi, L. (1982) Linguistic and social constraints in storytelling. Journal of Pragmatics, 6, 509-524.

Rumelhart, D. E. (1975) Notes on a schema for stories. In D. G. Bobrow and A. Collins (eds.) *Representations and Understanding: Studies in Cognitive Science*. New York, NY: Academic Press. pp. 211-235.

Rumelhart, D. E. (1977a) Understanding and summarizing brief stories. In Laberge, D, and Samuels, S. J. (eds.) *Basic processes in reading: perception and comprehension*. Hillsdale, N. J.: Erlbaum. 265-303.

Rumelhart, D. E. (1977b) Toward an interactive model of reading. In S. Dornič (Ed.) *Attention and performance*, 6, 573-603. Hillsdale, NJ. Erlbaum.

Rumelhart, D. E. (1980) Schemata: the building blocks of cognition. In R. J. Spiro, B. C. Bruce, and W. F. Brewer (eds.) *Theoretical Issues in Reading Comprehension*. Hillsdale, NJ: Erlbaum, 123-156.

Samuels, S. J., and Kamil, M. L. (1988) Models of the reading process. In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.) pp. 22-36.

Schank, R., and Abelson, R. (1977) *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Erlbaum.

Shavelson, R., Webb, N. & Burstein, L. (1986) Measurement of teaching. In M. Wittrock (Ed.) *Handbook of research on teaching* (pp. 50-91). New York: MacMillan.

Sinclair, J. McH., and Coulthard, R. M. (1975) *Towards an Analysis of Discourse*. Oxford: Oxford university Press.

Smith, F. (1971) *Understanding reading: a psycholinguistic analysis of reading and learning to read*. New York: Holt, Rinehart and Winston.

Stanovich, K. E. (1980) Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16, 32-71.

Stein, N. L. (1982) The definition of a story. *Journal of Pragmatics*, 6, 487-507.

Stein, N. L., and Glenn, C. G. (1979) An analysis of story comprehension in elementary school children. In R. O. Freedle (ed.) *New Directions in Discourse Processing*, 53-120. Norwood, N.J.: Ablex.

Steffensen, M. (1988) Changes in cohesion in the recall of native and foreign texts. In Carrell, P. L., Devine, J., and Eskey, D. E. (eds.) pp. 140-151.

Swales, J. M. (1990) *Genre Analysis. English in academic and research settings*. Cambridge: Cambridge University Press.

Tannen, D. (1979) What's in a frame? Surface evidence for underlying expectations. In R. O. Freedle (ed.) *New Directions in Discourse Processing*, 137-181. Norwood, N.J.: Ablex.

Thorndike, E. L. (1917) 'Reading as Reasoning: a Study of Mistakes in Paragraph Reading.' *Journal of Educational Psychology*, 8, 323-332.

Urquhart, S., and C. Weir (1998) *Reading in a Second Language: Process, Product and Practice*. London: Addison Wesley Longman Limited.

Venezky, R. L., and Calfee, R. C. (1970) The reading competency model. In Singer, H., and Ruddell, R. B. (eds.) *Theoretical Models and Processes of Reading*, 273-291, Newark, DE: International Reading Association.

Wallace, C. (1992) *Reading.* Oxford: Oxford University Press.

Weir, C., Huizhong, Y., and Yan, J. (2000) *An empirical investigation of the componentiality of L2 reading in English for academic purposes*. Studies in Language Testing, 12. Cambridge: Cambridge University Press.

Widdowson, H. G. (1978) *Teaching language as communication*. London: Oxford University Press.

Widdowson, H. G. (1983) *Learning Purpose and Language Use*. London: Oxford University Press.

Winter, E. O. (1977) A clause-relational approach to English texts: a study of some predictive lexical items in written discourse. *Instructional Science*, 6 (1), 1-92.

Yong-Won, L., and Yasuyo, S. (2007) *Cognitive Diagnosis Approaches in Language Assessment: An overview*. Paper presented at LTRC, June 9-12, 2007. Barcelona, Spain.

Appendices

Appendix A: The Reading Tasks and Answer Keys

<u>TASK 1</u>

You are going to read some advertisements. Match the advertisements (A-P) with the numbered sentences (1-10). There are five advertisements that you do not need to use. Write your answers in the boxes. There is an example at the beginning (0).

A EXPERIENCE the coast of Turkey on our super equipped 50ft Hinckley yacht up to 6 persons bareboat charter Tel. Finesse 01625 500241

B DINNER JAZZ by the Bob Moffatt Jazz Quartet. Live music for your wedding or social event. Tel 01524 66062 or 65720

C CHESTERGATE COUNTRY INTERIORS 88-90 Chestergate, Macclesfield. Tel 01625 430879 For quality hand waxed furniture traditionally constructed in antique style, from new wood.

D LA TAMA Small and select—intimate and inviting. The perfect place for a romantic meal and that special occasion. 23 Church Street, Ainsworth Village Tel. 01204 384020

E BRIAN LOOMES Specialist dealer with large stock of antique clocks. Longcase clocks a speciality. Calf Haugh Farmhouse, Pateley, North Yorks. Tel. 01423 711163

F THE ANSWER TO PROBLEM FEET. Hand crafted, made to measure shoes at affordable prices. Quality materials and finish. The Cordwainer Tel. 01942 609792

G REEVES DENTAL PRACTICE. The only BUPA accredited dentist in the Chorley area. 38, Park Road, Chorley. Tel. 01257 262152

H NEW AUTHORS publish your work. All subjects considered. Fiction, Non-Fiction, Biography, Religious, Poetry, Children's. Write or send your manuscript to MINERVA PRESS 2 Brompton Road, London SW7 3DQ

I COUNTY WATERWELLS LTD. Designers and installers of water wells and water systems. Bore hole drilling, water purification and filter systems. Tel. 01942 795137

J PENCIL PORTRAITS Unique gifts from Ł35. People, children, pets, houses and cars. Brian Phillips, The Studio, 14 Wellington Road, Bury, Lancashire, BL9 9BG.

K BANGS PREMIER SALON. Professional consultants in precision cutting, long hair, gents barbering, colour & perm. 149 Roe Lane, Southport. Tel. 01704 506966

L OUTDOOR GARDEN LIGHTING. Reproduction Victorian style lamp posts and tops, 3 sizes. Tops fit original posts. Catteral & Wood Ltd Tel. 01257 272192

M MARTIN HOBSON, advertising and commercial photography. For the best in the North West. Tel. Rochdale 01706 648737

N JOHN HAWORTH TELEVISION specialist dealers in quality television-video equipment. Competitive rates, free delivery-installation. 14abc Knowle Avenue, Blackpool Tel. 0800 0255445

O ABBEY EYEWEAR Designer spectacles with huge savings. Within grounds of Whalley Abbey, Whalley Tel. 01254 822062

P MALCOLM ECKTON Wedding and portrait photography. Treat someone special to a Hollywood make-over and portrait session. Ideal Christmas present. Malcolm Eckton, Studio, 18 Berry Lane, Longridge. Tel. 01772 786688

Write your answers here

0	*Julie wants to publish her book.*	**0**	H
1	Jack wants something old and valuable.	**1**	E
2	Jill wants a new pair of sandals.	**2**	F
3	Angela wants to eat out with her boyfriend.	**3**	D
4	Charles wants to go on an exotic trip.	**4**	A
5	Cathy has a toothache and wants a doctor.	**5**	G
6	Richard wants a band for his party.	**6**	B
7	Jane wants a new hairdo.	**7**	K
8	Peter wants home entertainment.	**8**	N
9	Jessica wants new glasses.	**9**	O
10	Roger wants pictures for his business.	**10**	M

TASK 2

You are going to read a magazine article about pandas. Some sentences are missing from the text. Choose the best sentence (A-G) for each gap (1-5) in the article and write its letter in the box. There is one extra sentence that you do not need to use. There is one example (0) at the beginning.

GIANT PANDA FACTS

Giant pandas are chubby mammals that live in a few remote mountainous regions in China. They have thick fur with bright black-and-white markings. **(0)** ____ The fur is water-repellent and helps keep a panda warm and dry in cold, wet weather.

(1) ____ Sometimes pandas eat other types of plants and occasionally they eat small mammals. But pandas usually eat only the stems, twigs, leaves, and fresh young shoots of the different types of bamboo. They especially like the tender shoots of young bamboo plants.

(2) ____ Full-grown pandas are close to 1.5 m tall when standing up and some grow as tall as 1.7 m. Males and females look alike, but females are a bit smaller than the males.

Pandas usually live alone. Each panda lives in an area that's about one or two miles (1.5 to 3 km) in diameter. **(3)** ____

In the spring, pandas search for a mate. They mark their territories with special scent glands to let other pandas know they are ready to mate. Once pandas mate, they separate and the females raise the young alone.

A new-born panda is only about the size of a hamster and weighs about 100 grams. **(4)** ____ And they have only a thin covering of hair. It takes a few weeks for the typical black-and-white markings to appear.

(5) ____ They learn how to find food, climb trees, and stay away from enemies.

[Ranger Rick's Nature Scope]

A   Some pandas live as long as 30 years and weigh as much as 117 kg.

B   Pandas are plant eaters and they feed mainly on a plant called bamboo.

C   In stormy weather, they sometimes try to find a cave or some other type of shelter.

D   Pandas are born without teeth and with their eyes closed.

E   Young pandas stay with their mothers for about a year and a half.

F   A panda's coat acts like a thick winter raincoat.

G   Although pandas will share part of their territory with other pandas, they don't usually get too close to each other.

Write your answers here:

0	**1**	**2**	**3**	**4**	**5**
F	B	A	G	D	E

TASK 3

You are going to read the first part of a newspaper article about gorillas in Uganda. Choose the most suitable heading from the list A - H for each part (1 - 6) of the article. There is one extra heading that you do not need to use. There is one example at the beginning (0). Write your answers in the boxes after the text.

Gorillas in Uganda's mist

(0)

A BLACK furry face stared out through the branches. Wide-eyed innocence tinged with mischief. After an hour and a half of hacking through forest, I was face to face with the mountain gorillas of Uganda. For 25 minutes I gazed, transfixed, hardly daring to breathe as two youngsters played out their daily lives, seemingly oblivious to the wonder-struck intruder.

(1)

Bwindi Impenetrable Forest, in the south-west, hides a remarkable secret. Designated a National Park in 1991, this magical, mist-shrouded area is home to roughly 300 mountain gorillas – half the world's population.

(2)

They are split into 23 groups, two of which are now habituated to human presence. The Mbare troop consists of 13 animals. The group was named after the hill – the word means rock in the local dialect – on which they were first spotted.

(3)

Six females and six young are led by the silverback male Ruhondezh – literally one who sleeps a lot. Ruhondezh, his back seemingly as wide as a bus, was magnificent. And it was clear that food, rather than sleep, was on his mind as we watched.

(4)

One minute, he munched contentedly on the vegetation while members of his family played in the branches above. The next, displaying his 8ft reach, he brought a huge branch crashing down to provide more sustenance.

(5)

Being so close to such impressive wild animals brings all your senses to life. In our passive, modern world, it is all too easy to lose touch with these primeval feelings. But in the heart of Africa, crouching just 15ft away, basic instincts rule. I felt a tremendous privilege at being allowed to share, even for a brief time, the lives of these gentle animals, which are on the edge of extinction.

(6)

To ensure their survival, the local people must feel there is some worth in keeping the gorillas. To such an end, the park authorities are currently engaged in revenue sharing. A percentage of the money raised from allowing tourists to view the gorillas is ploughed back into the community. In this way, it is hoped the gorillas will be seen as a source of income to be protected. But even so, the long-term survival of one of man's closest relatives hangs by a thread. Poaching is still one of the biggest dangers.

A How the gorilla population is organised

B Meeting the gorillas

C The leader of the group

D The location

E Appreciation of a unique experience

F The gorillas' reaction to seeing the author

G What is done to protect the gorillas

H What the leader of the group did

Write your answers here:

0	1	2	3	4	5	6
B	D	A	C	H	E	G

TASK 4

You are going to read a story about four friends. Eight sentences have been removed from the text. Choose from the sentences (A–I) the one which fits each gap (1-7). There is one extra sentence which you do not need to use. Write your answers in the boxes after the text. There is one example at the beginning (0).

'Being wet got us a train ban'

Jo Talbot and her three friends, all 13, expected the summer holiday to end with a bang — not a ban ...

'My three friends Jo Cole, Sara, Nicola and I all live in a small village outside Southampton. Last August we took the train into the city to go shopping for clothes one last time before starting the new term.

We got into Southampton at about 10am. **(0)** _____ No-one wanted the summer holidays to end, but it was as good a way as any to give them a send-off.

(1) _____ We weren't far from the station when the sky went black and there was a huge clap of thunder. We all shrieked and ran for cover, but the rain came down so hard it was like standing in a power shower. **(2)** _____

When we got to the station a train was waiting to leave, so I asked a guard if it was the one going to our local station. He looked at us and said, 'It is — but you're too wet to get on.' **(3)** _____

We were really fed up as we watched all the other passengers pull away, warm and dry. I couldn't believe they'd all avoided the rain, and got the feeling we were being picked on because we were kids. **(4)** _____

We sat around freezing cold, until the next train came along but strangely, we had no problem getting on that one. **(5)** _____

When I told my mum what had happened she was storming mad, and rang up South West Trains to ask them if they'd have treated an adult the same way. **(6)** _____ Customer services rang back later to say that the guard had been taken off duty while the company held an investigation.

It may not sound that bad, but the whole thing really spoiled our day. **(7)** _____ '

A   We'd have been happy to stand if they were worried we'd wreck the seats, but now we had to wait half an hour without even enough money for a cup of tea.

B   All my mates' mums wrote to the train company, asking if the same thing would have happened late at night, when we might have been put in real danger.

C   We were only caught in it for a minute but we were drenched — and were only wearing flimsy T-shirts and sweatshirts.

D   My friends and I were too shocked to argue, so we just let the train leave the station.

E   One thing is for sure, though, we're all taking umbrellas next time we go shopping.

F   Eventually we wandered back to catch the 2 pm train home.

G   We'd just got on the motorway when the car began to make a loud cracking noise.

H   On the journey back, I could hardly stop shaking with cold, and when I got back home I got straight into the bath to warm up.

I   We tramped around the shops buying loads of stuff and then went for a burger.

Write your answers here:

0	1	2	3	4	5	6	7
I	F	C	D	A	H	B	E

TASK 5

You are going to read a newspaper article about an unpleasant experience. Choose the most suitable heading from the list A-H for each part (1-6) of the article. There is an extra heading that you do not need to use. There is an example at the beginning (0). Write your answers in the boxes.

A   A trick - will it fail ?

B   An unexpected narrow escape

C   Two approaches to public use of office buildings

D   The best way to find shelter from the rain

E   An Englishman's home is his castle

F   A sudden obstacle

G   Possible short-cut ?

H   One problem made worse by another

Write your answers here:

0	1	2	3	4	5	6
H	G	F	A	B	C	E

Caught out in the rain

(0)

I was caught out the other day in a Manchester downpour (a much rarer event than is generally supposed, as the Met Office figures will readily confirm). Troubles never coming singly, the street down which I was hurrying to my appointment turned out to be blocked by some vast sewer reconstruction scheme. It looked as though I had no alternative but to retrace my steps and make a long detour. And I was getting wetter by the minute.

(1)

But suddenly salvation seemed at hand in the shape of a large office building that loomed up on my right hand side. Glancing through what was clearly a rear entrance, I could see across a wide lobby and the front entrance to the street on the far side – the very street I was trying to get to. There was something to be said for these new "prestige" office developments after all.

(2)

But not much. Completely blocking my path as I stepped through the swing doors into the lobby was a wide desk and behind it a middle-aged woman with a steely expression. "Can I help you, sir?" she said in a voice which suggested that that was the very last thing on her mind and that she knew very well what I was up to because I was the fiftieth person to use her lobby as a rat run that morning.

(3)

Now I may have been completely wrong in crediting her with such prescience but what followed suggests otherwise. "I have an appointment with Mr Henderson", I lied. "I think he's on the first floor." I waved my hand in the direction of the staircase and started off towards it. "Just a minute. I don't think we have a Mr Henderson." Without removing her eyes from my face for a second she picked up a house phone. "I'll ask Personnel," she said.

(4)

I was saved by the bell – the one on the phone on an adjoining desk. Putting down her own, she leaned over to answer it. Her eagle eye was off me and I was off towards the stairs and then to the door beyond and out into the street and the Manchester rain.

(5)

Let me be the first to say that that was a pretty silly way for a grown-up man to behave and it reflects no credit on me at all. But neither does it reflect any credit on those who administer ordinary commercial office buildings as though they housed both MI5 and 6 with the crown jewels lodged temporarily in the basement. In America such places are generally regarded as being in the public domain, with newspaper stands and snack bars. It may be hard on the flooring but most owners consider this easily outweighed by the good that accrues to the corporate image.

(6)

Here in Britain, I suppose, it's just the "Get off my land" attitude transferred from a rural to an urban setting. But it's sad to see this atavistic approach surviving even against its practitioner's own interests.

You are going to read the first part of a magazine article about animals. Some parts of the text are missing. Choose the best part from the list (A-J) for each gap (1-8) in the article and write its letter in the box. There is one extra part that you do not need to use. There is one example (0) at the beginning.

Animals under threat

- why should we worry about them?

For generations of children learning to read, their books have been filled with animals, from Babar the elephant to the *Jungle Stories* of Rudyard Kipling. But such creatures could become figures of nostalgia within a few years **(0)____**. The future is gloomy, according to Will Travers, director of Zoocheck, chairman of the protection group Elefriends and son of the campaigning conservationist Bill Travers. "Unless we act now, **(1)____**," Will warns.

His view is not exaggerated or alarmist: the fact is that **(2)____**. Sophisticated techniques, from test tube fertilisation to embryo freezing, can help to artificially 'save' endangered species, but what is the real point? Do we want to preserve tigers, for example, **(3)____** pacing up and down in a zoo? In a world **(4)____**, zoos are losing their popularity anyway. So wouldn't it be better **(5)____**? Or should we simply do nothing and accept extinction as Nature's way of ensuring 'the survival of the fittest'?

In 1839, the naturalist Charles Darwin first described evolution in his book *The Origin of Species by Means of Natural Selection.* David Attenborough explains Darwin's theory this way: "All individuals of the same species are not identical. In one clutch of eggs from a giant tortoise, for example, there will be some hatchlings which, **(6)____**, will develop longer necks than others. In times of drought, they will be able to reach the higher leaves **(7)____**, and so survive. Their brothers and sisters with shorter necks will starve and die. So those best suited to their surroundings will be 'selected' and able to transmit their characteristics to their offspring."

Evolution is a continual process – failure to adapt leads to extinction. In fact, of all the animals which have lived on earth, **(8)____**. "No species – and that includes the human race – has a lifespan of more than a few million years, which in geological terms is short," says zoologist Mark Carwardine.

A where the wonders of wildlife are available at the flick of a television switch

B because of their genetic make-up

C around 1,000 of our bird and animal species become extinct every year

D which haven't yet been eaten

E in 50 years' time elephants and rhino will inhabit only the echoing corridors of museums or the territory of a zoo

F if there are practical reasons

G as they rapidly die out

H to pour the time and money into preserving these animals in their natural habitats

I just so that our grandchildren can gape at them

J 95% have either evolved into something else or have become extinct

Write your answers here:

0	1	2	3	4	5	6	7	8
G	E	C	I	A	H	B	D	J

Appendix B: Sample Follow-up Questionnaire used in Study Two

Giant panda facts

	Very Easy					*Very Difficult*
Was this task easy or difficult?	1	2	3	4	5	6

Was this item easy or difficult?

Item 1	1	2	3	4	5	6
Item 2	1	2	3	4	5	6
Item 3	1	2	3	4	5	6
Item 4	1	2	3	4	5	6
Item 5	1	2	3	4	5	6

Being wet

	Very Easy					*Very Difficult*
Was this task easy or difficult?	1	2	3	4	5	6

Was this item easy or difficult?

Item 1	1	2	3	4	5	6
Item 2	1	2	3	4	5	6
Item 3	1	2	3	4	5	6
Item 4	1	2	3	4	5	6
Item 5	1	2	3	4	5	6
Item 6	1	2	3	4	5	6
Item 7	1	2	3	4	5	6

Caught out in the rain

	Very Easy					*Very Difficult*
Was this task easy or difficult?	1	2	3	4	5	6

Was this item easy or difficult?

Item 1	1	2	3	4	5	6
Item 2	1	2	3	4	5	6
Item 3	1	2	3	4	5	6
Item 4	1	2	3	4	5	6
Item 5	1	2	3	4	5	6
Item 6	1	2	3	4	5	6

What did you find most difficult answering the items?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

Appendix C: Sample transcriptions and notes

Transcript No 1
Protocol produced by High-level Student 1 - **HS1**
Task: **Giant panda facts /** Matching sentences to gaps in text

(After reading the title of the text, she starts reading the text itself. Reads silently for 18 seconds.)

R: Remember to keep saying aloud what you think.

Now I'm completing the text. (Refers to Paragraph 0, which is gapped to provide the Example.) / I'm looking at F / that what belongs to the . the text / that is how the text continues and the . . / their thick fur is worth much / much . / that is expensive and so . . they make thick winter coats from it . / rather cruel / . . . their fur is . perhaps resistant . to water . . and against weather / that is resists those conditions as well . / and then here something is missing / (Reads into Paragraph 1 silently for 6 seconds and then turns to the Options.) I'll look at what possibilities there are . . . / I don't think it would be about their age if here it's about something . . / it's about their meals / about what they eat . / and the . B talks first about this / (Looks through the rest of the Options.) . . The third one is about weather . / then about the birth of pandas / that what a new-born looks like . . . / then again about young pandas . / We've already written in F / I cross that out . . . / (Looks at the last option, Option G, in the list.) I don't think they start a paragraph with '*although*' but who knows . / and that this isn't about meals either / I think B will fit in here . . / That is they are plant-eaters . / mainly . they eat bamboo . and also other plants / this is good here, I think / . . the '*mammal*' / I can't think of what it means. / (Reads silently for 20 seconds.) I don't understand this paragraph here / what it's about / . . The next one [Paragraph 2] is about adult pandas but here again something is missing before it. / (Checks the Options for 12 seconds.) There must be something about .?. / . I cross out B, that's done / (Turns back to the text) . . . . here after all . several . would fit so I'm going on . something will then get here on the basis of elimination / (Reads through Paragraph 3 and checks options for 20 seconds.) I think it's G that will be good for the next one . / a comparison that . / where they live and how / if in groups or alone / (Reads silently for 18 seconds.) They mark their . territories / . . . it will be about reproduction here . . . . / They talk about the new-born panda in this one (Refers to Paragraph 4) and this was already mentioned at one of the . letters / (Pauses for 11 seconds.)

R: Remember to keep talking.

Now I'm looking for the one which . fits best here . in the text / that which one is about how pandas are born / and the . the D . describes how . / without teeth and . eyes / or with closed eyes . / I think this will be the suitable one here . . / this is also a description . . . . / They conceal themselves with their fur . / oh, no, discover / (Goes on to read the last paragraph, Paragraph 5) . . . . Here it's about nutrition . and how they stay . . how they stay away . from enemies . . . / I'm looking for that that . which one may be suitable here . . / I think . E describes best that . how they do these that . I see here in the paragraph . with its parents . . / Now I'll return to what I haven't filled in yet. / (Goes

back to Paragraph 2.) . . About the adult panda . full-grown . . . / There are three more [options] left . / maybe . somewhere I didn't write in or didn't cross out / (Pauses for 15 seconds.) They write that there will be one extra . sentence and . / for two three four six . places . seven . / I don't know / then I've done something in the wrong way or I can't find . . . / I think for 2, A will be the best . / oh, yes, I didn't cross out G / then I've got what I didn't find / (Crosses out Option G.) and A at 2 / then C remains to be an exception. / Now I'll read through if this makes sense. /

R: Why do you think A goes with 2?

Because . the paragraph is about . adult . the adult p full-grown pandas and . and here (Refers to Option A.) it talks about their age and weight, while in the others, that is, above all in C . . in C it talks about a kind of protection, the weather and . some cage . . / to protect / about a shelter . . . / while here, at 2 . the same way as in A, first it writes down what / how old it can be / how . heavy / and then the paragraph continues that how tall it is . / and compares the . female and the . the male pandas . / Well, I'll look at if it makes sense . what I've written. / (Checks all her answers, starting with the Example item.) F will be good . / They said that. / (Pauses for 14 seconds.) B, I think, connects well. / . . . After A, the next sentence fits. (Pauses for 14 seconds.) A comparison within the paragraph at the third [gap/item]. / (Pauses for 33 seconds.) I don't understand what exactly it writes about the new-born, but it's about them / and here about how they are born . in D . / so this fits in here. / (Pauses for 12 seconds.) It's interesting that it takes a few weeks for them . to have black and white hair / . . . Then it writes about young pandas, again . / so E fits here . / Then I write it in the frame because . they will take that into account. / So 1 is B, 2 is A, 3 is G, 4 is D, 5 is E.

R: You would finish it here at an exam?

Yes.

Notes on task processing

High-level student - **HS1**
Task: **Giant panda facts** / Matching sentences to gaps in text
Time spent completing the task: 12 minutes

General notes

1 Reads the title of the text.
2 Reads the text silently paragraph by paragraph and summarizes the information she finds important in each paragraph.
3 Tries to identify the topic, understand the main ideas in each paragraph, not worrying much about unknown words, or even sentences that she does not (fully) understand.
4 Completes the Example Item in Paragraph 0, that is, reads through the paragraph and checks how the sentence taken out of the paragraph to provide an example is connected to the sentences before and after the gap.
5 Answers the items as she is reading the text paragraph by paragraph. That is, she reads Paragraph 1 and answer Item 1, before going on to read the next paragraph. Then reads Paragraph 2 and attempts to give an answer to Item 2, before reading the next paragraph, and so on.
6 Reads through all options (missing sentences) when, after reading the first paragraph of the text with the Example item, she reaches the first numbered gap to be completed (Item 1).
7 While responding to the items, she pays attention to crossing out an option from the list as soon as she has used it as an answer.
8 When she has completed all items on the task, she goes back to check if the sentences she inserted in different sections of the text indeed fit in with what comes before and after the inserted sentence.
9 She gives a correct answer to all five items on the task. In the follow-up questionnaire, she assesses the task as 'very easy', rating it "1" on a 1-6 scale.

Notes on responding to the task, item by item:

<u>Item 1</u>
She identifies the topic of the paragraph and gives a correct answer to the item easily, on reading the text for the first time, although there are some words in the paragraph that are unfamiliar to her (e.g., '*mammal*') and so she does not understand all the details in the text.

<u>Item 2</u>
She gets the item right, but this is the only item that, on reading the paragraph, she leaves open to be answered at the end when she has already responded to all the other items on the task. Her verbal report shows that, on reading through the paragraph for the first time, she considered several options as possible answers to the item, regardless of the fact that, as is also clear from her report, she understood the content of the paragraph in detail. After responding to all the other items, when there remained only three (in fact, only two) options from which to choose, she selected the correct answer very easily, comparing the content of the two options both against each other and the content of the paragraph.

Item 3
She gives a correct answer to the item relatively easily, on the first reading. In the follow-up questionnaire she assesses the item, along with Item 2, as slightly more difficult than the other three items on the task. In the case of this item, one reason for this might be that, as her report suggests, she is likely to have had some difficulties understanding in detail the paragraph that immediately follows the item.

Item 4
She gets the item right, identifying the same topic in the text and the correct answer very easily. There is, though, a sentence in the paragraph that she does not seem to understand.

Item 5
She gives a correct answer to the item, identifying the relationship between the paragraph and the correct option easily.

In the follow-up questionnaire, she assesses Items 2 and 3 as the most difficult items on the task, rating both "3" on a 1-6 scale. One reason she mentions why she marked certain items more difficult than others is that, as she says, 'for example, in the paragraph about adult pandas [Item 2], it wasn't clear right away what exactly fits there'.

Transcript No 2
Protocol produced by High-level Student 1 **- HS1**
Task: **Caught out in the rain** / Matching headings to text

(Reads the instructions to the task and then starts reading the Options / paragraph headings from A to H.)

(Reads Option A) A brick [misreading the word '*trick*'] . / (Reads Option B) An unexpected . . pretty near escape / I don't know what it could mean. It'll become clear from the text. (Reads Option C) . . . . it's about office building . (Looks through the three Options she has read so far.) / In the first it's brick / in the second . / I don't know . / then office building . /

R: Could you speak up a little?

Yes, so in the first one it's brick, then some kind of unexpected event, then . office building . or something similar / (Reads Option D) . '*find shelter*' . it's that how we should find . shelter from the rain / Reads Option E) . . *gentleman* . . '*castle*' . . / his castle . / This will also become clear from the text . what happens to the English gentleman / (Reads Option F) . . an unexpected obstacle / the '*obstacle*' / I'm not sure about it. (Reads Option G) '*Possible short-cut*' / . short circuit (rising intonation) / (Reads the last option, Option H, given as an Example.) '*One problem made worse by another*' . . / Well, let's see . perhaps it becomes clear after all what this is / perhaps / I'll see / *One problem made worse by another* . . / I read the text. /

(Turns to the text and starts reading Paragraph 0.) (Reads silently for 11 seconds.) Manchester reminds me of football. (Reads silently for 15 seconds.) Trouble never comes singly. / (Reads silently for 56 seconds.) Well, I didn't understand much from this here. / They got drenched in a minute. / I think it's good that it's not me who has to write it in . which title is needed for it. / I'd rather go on . so that time is not taken away by this.

(Starts reading Paragraph 1.) (Reads for 12 seconds) '*Salvation*' / I don't know what it means / not even familiar (Pause – 12 seconds) / on the right hand side / an *office building* . / In [Option] C, it talks about *office building* but who knows maybe it's tricky (Pause – 19 seconds) / '*rear entrance*' / some kind of entrance / (Pause – 18 seconds) The first entrance . . . looked onto . the street . / it's on the far side . . / This is interesting. / '*the very street*' . . . / just in that direction / which I'm sure is odd at first . . / the street . . where he wanted to go . / that . (Pauses for 24 seconds, during which she seems to be reading the last sentence of the paragraph.) Maybe I should re-start from the beginning because I don't understand very much. / . . . . '*"prestige" office developments*' . . / Well, in any case . I skim through the text once.

(Starts reading Paragraph 2.) . . . . blocked his way / . . '*swing doors*' that's . swing door . perhaps / Swing is the . / it appears in dance as well. / (Pause – 17 seconds) '*a wide desk*' / (She reads the rest of the sentence in a very quiet voice.) '*and behind it a middle-aged woman with a steely expression*' . / *steely expression* . / at the woman . / It was at the entrance exam yesterday . they asked about the iron*lady* who she is / I have no idea and here is a woman who whose reflection is steely . / Goes on to read the rest

of the paragraph.) '*Can I help you, sir?*' (Reads silently for 25 seconds.) / '*I was up to*' / this . this is a kind of expression . . . / I still don't know what '*lobby*' means and this has already occurred many times . . / And I am the 50th person . who lays claim to . her services / I don't know what this is intended to say . / rat run / (Smiles.) I'm not familiar with this . in this form . / I guess it's the morning rush or . something similar . . / I look at what . what it is that may fit here this paragraph but . I don't have much chance . / (Turns to the Options and checks Option F against the paragraph.) An unexpected/sudden obstacle, perhaps . . / Well, yes, here after all the . / it writes that something blocked his way . . as soon as he stepped through the . swing door . . . / I write G here, perhaps that will be the good one . / I marked it / [Although she explains why she thinks Option F is the correct answer to the item, when marking her answer, she selects, by chance, the letter of another option, which comes right after Option F in the list of options] (Returns to the text and starts reading Paragraph 3.)

. . . . *credit* / that that . I've heard that only with . *credit card* / so far / . . and here it's used in connection with a person / as a verb / . '*crediting her*' (Pause – 19 seconds) / the following '*suggests . otherwise*' / something else . it suggested or . . I don't know what it means / '*I have an appointment with Mr Henderson*' / . So, he lied. (Reads on for 18 seconds.) And was caught out / there is no Mr Henderson in the house. / (Reads the rest of the paragraph for 15 seconds.) / .?. / . So, this didn't work. / . No. / . Don't understand / I mean the . / the woman / that that what she wants . / that / I / I don't know what this man wants but . he wants to get in this house and he didn't succeed . . . / I look at if anything refers to this . among the answers. (Looks through the Options for 10 seconds.) (First checks Option B against the item.) An unexpected pretty near escape (rising intonation) / perhaps. / (Pause – 24 seconds) Here . at C . / it's about public use / the office building . also has it . / and well, after all, here also he wants to get in if I know it well . / It troubles [me] . . / I don't think D is good because if he only wanted to find shelter from the rain then why would he want to get into the building. / Perhaps only because . . because he still has to say something . . . / I'd rather go on with the text / maybe something . becomes clear.

(Starts reading Paragraph 4.) '*saved by the bell*' (Reads silently for 19 seconds.) / She put down her own and / '*leaned over to answer it*' / . . she's got eagle eyes / '*eagle eye*' / . . why isn't it in plural / (Pause – 11 seconds) It was on me . off me . . / I don't understand / (Pause – 16 second) Oh, I see / that he was shown the door . . . first towards the stairs and then . . . '*the door beyond*' . '*out into the street*' and then he was in the Manchester rain. / . Well . . . . (Checking the Options, she realizes the mistake she made when marking the heading for Paragraph 2.) I wrote a wrong answer for [Item] 2 / that's F . not G. / . . . . It's possible that [Option] A / here / has remained . I don't know / . . . I don't think this is the best . method or way for him to . to find shelter . I don't think D is good. (Checks Option C.) '*Two approaches*' . . / Here it's about two . approaches / . while here he was thrown out. (Checks Option B.) '*unexpected narrow escape*' / (Pause – 11 seconds) / .?. / (Without deciding on her answer to Item 4, she returns to the text.)

(Starts reading the first sentence of Paragraph 5.) '*Let me be the first*' . / this is quite good an expression / '*to say that that was a pretty silly way*' . . . / He behaved in a silly way. (Pause – 12 seconds) '*credit*' / can that be trust or something similar . / his trustworthiness / his credence / well . if *credit card* is credit card then . perhaps what he lost was his credit / trustworthiness . / doesn't matter / (Reads the second sentence of the

paragraph silently for 33 seconds.) '*both MI5 . and 6*' / . . . and the crown jewels . / '*temporarily . . in the basement*' / temporarily in the . basement . . . / I don't understand this . . . / Neither does he consider better those who . '*administer ordinary*' (Pause – 14 seconds) / '*commercial*' is a kind of . . / it's got to do with . trading / '*office building . as though . they housed both*' . . . / (Gives up trying to understand the meaning of the second sentence and goes on to the third one.) In America / sure enough / . . . these . buildings . . are public . / public *domain* / that can be a user because they use a kind of *domain* name at the / the email addresses / . . . '*newspaper stands and snack bars*' / . . with newspaper stands . oh, yes and with buffets / then . they also have attendants / (Pauses for 59 seconds.) I look at if perhaps C fits here. / Well, the two approaches / I don't know what '*approaches*' means / that's the trouble . / but it's a kind of . / if it meant attitude then it would fit here . / the English and the American / so . for the time being I write it in . . / Then, I'm going on. /

(Starts reading the last paragraph, Paragraph 6.) (Reads silently for 13 seconds.) Get off my land. / '*Get off my land*' / (Pause – 37 seconds) This is a summary. / . . the Englishman's house is his castle / This this E . seems quite good here / . instead of Az én házam az én váram [My house is my castle] / . . . For the time being . let it be E . / and then I start it from the beginning. /

(Returns to the beginning of the text and starts reading through it for the second time. Attempts to synthetize what she has understood in different sections of the text and, at the same time, finalize her answers and identify some more matches between paragraphs and headings.)

Well, he got drenched / then . . . he noticed the . *office building* . on the right hand side / . . . . and there he clearly saw an . entrance / '*across a wide lobby*' . / the *wide lobby* / I still don't know what that is / '*and the front entrance*' . . . . '*to the street on the far side*' / on the far side in exactly the street . that he wanted to get to / perhaps he got there . . / (Reads the last sentence of Paragraph 1 in a very quiet voice.) '*There was something to be said for these new "prestige" office developments after all.*' / . . . passive structure / . . . prestige offices' . development . / (Attempts to finalize her answer to Item 1.) I cross out E, I've already written it in and I've also written in C / (Pause – 12 seconds) Maybe this the best way for him to find shelter (rising intonation) . / Then / but then D would be good . . / Let it be D . . / Then he doesn't yet know that they will kick him out. / (Goes on to Paragraph 2) . . Then . . then he continues . / (Checks Option F, the correct answer to Item 2.) The unexpected/sudden obstacle comes here . . / He tries to stay in the building / he doesn't want to go further in / he just wants to stay in. . / (Goes on to Paragraph 3.) This is why . he wants to get to the staircase. / But there he lies. (Pauses for 34 seconds.)

(Attempts to decide on her answers to the two items she has not yet responded to, Items 3 and 4.) Now, I have three options that is titles left [Options A, B, G]. for two places. / (Checks Option A) . . . I don't think there is anything about brick . in this article. / (Pause – 12 seconds) (Checks Option G) '*short-cut*' / that's . / now that's either short circuit / but it is also possible that that a kind of . shorter route . / yes it rings a bell that . at travelling we . we used that . / to cut off the way / (Pause – 31 seconds) (Tries to clarify the meaning of Paragraphs 3 and 4.) Here's the staircase and / well, he wants to get to the staircase / in the other (End of tape) / .?. / In this one, he still wants to get in while here . he's shown the door. / (Checks Options B and G.) '*Possible short-cut?*' /

'*unexpected narrow escape*' / (Pause – 22 seconds) I may have done something in the wrong way. / (Checks her earlier answers to Items 5 and 6.) Well, C / that that / that seems to be good. That . I'm completely . sure about that / this is a comparison . two approaches / Then a summary / [Paragraph] 6 is E . . . / Then . . . I think I exclude [Option] A. / Such a dangerous situation is not mentioned here . . . unless I completely misunderstand '*trick*'. / (Pauses for 11 seconds.) Well, who knows. /

(Checks her earlier [correct] answer to Item 2 [Option F] and thinks about changing it to an incorrect one.) The unexpected/sudden obstacle / that . . . / well, it's possible that it doesn't . . / it doesn't refer to the '*wide desk*' . or . the iron *lady* . . but that they're on the phone and they put him out. (Pauses for 11 seconds.) (Checks her answer given to Item 1.) '*The best way to find shelter from the rain*' / this seems to be good for 1 / at that point he can still be optimistic / . . . Then there remain three options [B, G, F] for three places [Items 2, 3, 4]. / . It would only need to be decided for which. / What else are there / let's see / (Looks through, mainly, Paragraphs 2, 3 and 4 for 27 seconds.)

Difficult people . . / His attitude to the woman . was completely wrong. I don't know what this is meant to suggest. / . . . . Oh, yeah that it occurred to him that he may have judged her wrongly . perhaps / . but what followed this / that . . . that advises . otherwise or . something like this . . . / So, he lied. / I don't understand why he lied. / . . . He trusted he would be let in . so he lied / '*I think he's on the first floor*' / . and in the end he's not let in / (Pauses for 22 seconds.) (Makes an unsuccessful attempt to finalize her answers to Items 3 and 4, checking Options B and G against Paragraphs 3 and 4.) For 4, I can't write '*An unexpected narrow escape*' [Option B] 'cause there / because there . there . . there he's put out / that is there . nothing happens / there only an unexpected thing happens but . nothing happens that nearly happens to him / something concrete. . / In [Paragraph] 3, he nearly . escapes. / After all, he may think that he nearly escapes. / . . . But this could be the / . . . . / could go to 3 as well / while '*short-cut*' [Option G], if it was the shortest way . . / that's also a good . option / well, no / no, it's rather this one that'll be 3, I think / . I write it in / G is 3 . / and 4 is . / well, this is not good / (Pauses for 15 seconds.) [Item] 2 . / maybe 2 is wrong (rising intonation) / (Pauses for 13 seconds.)

(Attempts to clarify her answers to Items 2, 3 and 4.) If I write F for 4 / as unexpected/sudden obstacle . . / because . he was stopped . . / then for 2 / perhaps he's still hoping and / perhaps it's G that is good there after all / which I wrote in by chance . / Let's see if this is possible. / Short-cut [Option G] . . and if for 3, I write in '*unexpected narrow escape*' [Option B] / (Pauses for 20 seconds.) Well . I'm not sure it's good. (Pauses for 11 seconds.) I'm already not sure about the meaning of '*trick*', either . . . / If it's brick then that has no business here . .

(Finalizes her answers.) I'm sure about C and E . / Then I write that in. / C . is 5, E is 6 . . . / '*The best way to find shelter*' . . . / Also, D fits 1 quite well / I also write that in . . . . / Then [Option] A / I've excluded that one . / '*An unexpected narrow escape*' [Option B] / '*Sudden obstacle*' [Option F] / and '*Possible short-cut?*' [Option G] . / Well . three for three places / . that can be in six different ways (Smiles.) . / I don't have much chance but then . . . / I write in F for 4. / '*Possible short-cut?*' [Option G] surely doesn't fit there. / Neither does the pretty near escape [Option B], I think. / (Re-checks, by repeating Option F, her answer to Item 4.) '*Sudden obstacle*' / unexpected/sudden obstacle . . . . / Then there remain two for two places / that's only . four . / oh, no, only

two different ways. / '*unexpected narrow escape*' / '*Possible short-cut?*' . . . / (Considers Option G.) Well, this is an interrogative . / *possible . possibility* / this seems quite hoping so let this be in the first place / (Writes G in the answer box for Item 2.) Then let this come to place No 2 . / G and then for [Item] 3 . there remains B. / I can't find out anything better than this . .

R: So now you've finished?

Yes.

Notes on task processing

High-level student - **HS1**
Task: **Caught out in the rain** / Matching headings to text
Time spent completing the task: 35 minutes

General notes

1   Reads the instructions to the task.
2   Reads the options/paragraph headings before she starts reading the text.
3   Does not skim through the text before starting to read it in detail paragraph by paragraph.
4   Tries to understand the main ideas in each paragraph, regardless of the fact that each paragraph contains a number of difficult, low-frequency vocabulary items that she is not familiar with.
5   Does not give up trying to make sense of the text even if there are sentences whose meaning she does not at all understand.
6   Makes efforts to understand the relationship between individual paragraphs of the text (with special regard to the first 4-5 paragraphs), trying to understand the meaning of the text not only at the level of the paragraph but also at the level of the text as whole.
7   Often makes comments on what she is reading (although, in some cases, they are very difficult to understand because of her quiet voice).
8   Sometimes she re-reads or repeats words or phrases that she likes in the text.
9   Tries to answer the items, find a suitable heading for each paragraph while she is reading the text for the first time. When she has read through the text once, she goes back to re-read those sections for which she could not identify a suitable heading on reading the text for the first time.
10 To make sure she has selected the correct answer, she almost always checks all (remaining) options against the item to be completed.
11 Either does not understand or overlooks important details in Paragraph 0. As the paragraph presents the beginning of a 5-paragraph long narrative within the 7-paragraph long text, this makes it very difficult (in fact, impossible) for her to fully understand events of the narrative in, and find a suitable heading for, Paragraphs 1-4.

Notes on responding to the task, item by item

<u>Item 1</u>
She does not fully understand the main idea in the paragraph and gives an incorrect answer to the item. One of the main reasons for this is that there is low-frequency vocabulary, or vocabulary she is not familiar with, in the crucial information in the paragraph (e.g., '*glancing through*', '*rear entrance*', '*lobby*'). Perhaps more importantly, in order to understand the main idea in the paragraph, i.e., in Paragraph 1, one needs to understand the main idea, as well as certain details, in the preceding paragraph, i.e., Paragraph 0. Her verbal report suggests that she understood the main idea only in part and, besides, either did not understand or did not pay due attention to a crucial detail included in a key sentence in the preceding paragraph ('*It looked as though I had no alternative but to retrace my steps and make a long detour.*'), which is

likely to have contributed, to a great extent, to her failure to identify the main idea in Paragraph 1.

Another reason why she got the item wrong might be that the distractor is a very plausible answer to this item, even worse, only to this item and none of the others. It can only be excluded as an answer to the item if one understands, apart from the main idea, also the meaning of the above-cited key sentence in the preceding paragraph.

One source of difficulty in identifying the main idea and understanding details in the preceding paragraph is likely to be the relatively high number of difficult, low-frequency vocabulary items (e.g., '*readily confirm*', '*vast sewer reconstruction scheme*', '*retrace*', '*detour*') and some long and complex grammatical structures used in the paragraph.

Note: The sentence cited above can be considered to be a key sentence in Paragraph 0 insofar as if we take it out of the text, then the heading selected by the student as an answer to Item 1 might as well fit Paragraph 1.

Item 2
She gives an incorrect answer to the item. Although she is uncertain about some details, she understands the main idea in the paragraph. Neither does she have difficulty in understanding the meaning of the correct answer (Option F, *A sudden obstacle*). In fact, at one stage in the process of responding to the task, she selected and marked the correct option. Her verbal report shows that her first choice of answer was not a guess, but was based on her recognition of the relationship between the information she understood from the paragraph and the correct answer.

The main reason why, at a later stage, she reconsiders the answer she selected earlier and changes it to an incorrect one can be traced back to her misunderstanding of the main idea in another paragraph, Paragraph 4. Due to her failure to understand the main idea in Paragraph 4, she is unable to identify the correct answer to that item and, ultimately, finds the correct answer to this item to be a (more) suitable answer to Item 4 (which obviously means that she gets both items wrong). However, it is also apparent from her report that there is much uncertainty behind her final selection of the answers to both this item and item 4. What could have helped her clear up uncertainties around the item is, as in the case of Item 1, the information she did not understand in Paragraph 0. It would seem reasonable to suppose that a correct response to this item requires, apart from understanding the gist of the paragraph, an understanding of the content of Paragraph 4, and some details in Paragraph 0, as well.

Item 3
She understands the main idea in the paragraph, yet she gives an incorrect answer to the item. The apparent reason for this is that she misreads the word '*trick*' for '*brick*' in Option A, the correct answer to the item (*A trick – will it fail*?), which obviously makes no sense in either this or the other paragraphs of the text, therefore she excludes the option as a possible answer to any of the items on the task, including Item 3. Consideration of the correct answer to the item as a distractor, i.e., as an option that is not needed, leads to total confusion when she tries to identify a suitable heading not only for Paragraph 3, but also Paragraphs 2 and 4. Her verbal report clearly shows that

her misreading of the word '*trick*' in Option A contributes to her failure to give a correct answer not only to this item but also Items 2 and 4.

Note: A low-level student, whose vocabulary does not typically include the word '*brick*', would not be able to make the same mistake, with all its consequences, as this high-level student when responding to the item. To this extent, for a lower-level student, it might be easier to give a correct answer to this item (on the basis of the relatively easily recognisable relationship between the words '*lie*', in the text, and '*trick*', in the correct answer) than for a higher-level student.

Item 4

She gets the item wrong, as she misunderstands the main idea in the paragraph. The source of her difficulties in selecting the correct answer seems to be threefold. First, she has difficulties in understanding, or applying her knowledge of, the structure '*be off*', which carries crucial information in the paragraph ('*her eagle eye was off me*'). Second, there is no apparent sign in her verbal report of considering, or attributing any particular importance to, the first sentence of the paragraph ('*I was saved by the bell*'), which again includes crucial information for a correct answer. Third, lack of understanding details in the introductory paragraph of the text (Paragraph 0) mentioned earlier is likely to make it more difficult to understand the gist of the story as a whole and, accordingly, to fully understand the main idea and select a heading for, in fact, any of the four paragraphs that are involved in presenting details of the story (Paragraphs 1-4).

Item 5

She gives a correct answer to the item. Although she does not understand many details in the paragraph, she understands the main idea and, despite the relatively high number of low-frequency words and occasionally rather long and/or complex grammatical structures in the paragraph, is able to recognize the relationship between the paragraph and the correct answer relatively easily, on reading the text for the first time.

Item 6

She answers the item correctly. She has no difficulty in identifying the relationship between the paragraph and the correct answer. She selects the correct answer relying exclusively on her knowledge or understanding of the sayings '*Get off my land*', in the paragraph, and '*An Englishman's home is his castle*', in the correct answer.

In the follow-up questionnaire, she assesses Items 2 , 3 and 4 as 'very difficult' (with a rating "6" in each case). As in the follow-up session she explains, sometimes, as in the case of Item 2 in the Pandas task, she finds an item at the beginning of the task more difficult than an item around the end, because at the beginning there are a lot of options to choose from, which means she has to check all options before choosing the answer, while around the end of the task, there are fewer options so it might be easier to decide on the correct answer to those items. However, in this task, although at the end she had only three options from which to choose, it was not any easier for her to decide which option fits where in the text.

She also reports that she 'didn't understand what the text was about', 'there were many unknown words, or partially unknown words' that she knows she has learnt but has 'no idea what they mean' or she has 'only vague memories about them'.

Transcript No 3
Protocol produced by a low/middle-level student - **LMS**
Task: **Julie wants (Advertisements)** / Matching topic sentences to text

(Reads the instruction to the task. However, she gets confused about what the task requires her to do. She thinks that the sentences 1-10 [the ten items] need to be put in the correct order. It takes her about two-three minutes to notice, with some help from the researcher, that the task is longer than the page on which she can see the list of sentences. Realizing that the task is arranged on two pages, she also understands that she is expected to find matches between the sentences on the first page and the advertisements on the facing page. She counts the options to make sure that the five extra advertisements that she read about in the rubric are indeed among them.)

(She starts the task with the Example item.) (Reads silently for 30 seconds.)

R: Remember to keep saying aloud what you think.

Oh, yes. Now I've read through these. In [Item] 1, he wants something old but I don't know what. In 2, new sandals, in 3 .. she wants to go to a restaurant or somewhere with her friend, in 4, he wants to travel, wants to go on an exotic trip, 5 has a toothache and needs a doctor, 6 . . would like a kind of band at the party / at his party, 7 . needs a new . I don't know what, 8 . . . wants a kind of home . something, 9 – glasses, while 10 . . picture for the . company / . of the company / I don't know.

(Starts reading the options/advertisements.) . . . . In [Option] A, they want to sell something . something Turkish I don't know what / for 50 Ft . / of 6 persons (Pause – 14 seconds) / a what (rising intonation) (Pause – 15 seconds) I don't know, let's go on. / . . . Well, [Option] B is about a music band and then there was such a sentence / 6 / I think that will be the one / (Reads Option C for 17 seconds) some furniture . shop or something like this . . . . and kind of antique things / so it might as well be [Item] 1 because there was something old in that as well but I don't know what that is / or there was a kind of / for flat [Item 8] / so this might as well go with two [items] . I think / (Reads Option D.) . . . . this is some restaurant and a romantic meal and there was something like she wants to go for a meal with her friend / that's 3 and this one is D and so these will match / . I cross out D / (Reads Option E for 14 seconds.) This is some . clocks . . . the maker of . antique . clocks / I don't know this yet / (Reads Option F for 16 seconds.) This is about shoes and kind of problems with feet and there was one that would like new sandals . / that's [Item] 2 / then this will fit / . . . [Option] G is a dentist and there was one with a toothache . / that's [Item] 5 then that's G / . . [Option] H was the Example / (Reads Option I for 18 seconds.) [Option] I doesn't fit any of them, I think . . / it's about kind of waters / all kinds of water things and there wasn't such a sentence in the other [list] / (Reads Option J for 11 seconds.) A kind of . drawing studio / something like this / . . . he would like a kind of picture / that is / this is a maker of pictures and there was here that he would like a picture . . for the company and then that's . J for [Item] 10 / (Reads Option K for 11 seconds.) This is a hair salon and there was that she'd like a new . hair . something / well, then, that will be good I think / (Reads Option L for 16 seconds.) This is kind of outdoor . / it sells outdoor garden lamps but . . there wasn't such a thing / (Reads Option M) . . . . This is a photographer . / don't know this yet / (Reads Option N for 18 seconds.) TV and video . equipment / he

makes TV and video equipment / . . . . well this isn't good / I don't think / (Reads Option O.) . . . . I don't know / I don't understand what / what it wants / (Pause – 16 seconds) but in fact here it's kind of eye . wear . / well, it must be something to do with glasses and there was one about glasses . and that / no others at all would fit that / (Reads the last option, Option P) . . . wedding . portrait . . . / he makes kind of wedding and portrait photos . . / This doesn't fit any of them / left out / Then I'll look at what is left out . here /

(Returns to the list of items to look at those three, specifically, Items 1, 4 and 8, that she did not answer while reading through the options.) I still don't understand [Item] 1, that is, there are problems there . . / I understand that he would like something that is old and . I don't know what kind . . / In [Item] 4, he'd like to travel . / well, one that's got to do with travelling might be the . the first one [Option A] at most, because there it mentions Turkey . / 6 persons / well, I'd probably put that one to that / (Looks at Item 8) . . home equipment . no . or . . . . probably / that is / it's probably [Option] C, because kind of furnit / there / it's that in which there is kind of furniture and . antique and new and all kinds . and / well, that is considered to be home . equipment, indeed.

(Returns to Item 1, still unanswered.) And then [Item] 1. (Pauses for 18 seconds.) I read through what is left out. / . . [Option] I is surely not good for it because that's / that's about something completely different / . . . [Option] L, too . is about something else / . . the photographer one isn't either / . . . . I think it's [Option] E because something / he'd like something . old and it's that in which there is that / . kind of antique . clocks . / and that's the only one that . counts as old. / . yes / . and then it's done / . Now if there is time for it, I'd look through them if they are good.

R: There is time for it, just go on if you like.

Then [Item] 1 is E, that's OK. [Item] 2 is probably good because none of the others mention footwear so that's probably good. In [Item] 3, a restaurant (Pause – 12 seconds) / there is again only one restaurant so that's probably the one. [Item] 4 / I'm not sure about that but . . . I can't find anything better for that / for the time being. [Item] 5 is good for sure because that / the dentist was very straightforward. / . . . . [Item] 6 too, / there was only one of that, too. / . . . [Item] 7 is like this, too / there was only one about the hairdresser's. / I'm not sure about [Item] 8, either / . . . and then 9 and 10 are again straightforward / so then that's all.

Notes on task processing

Low/Middle-level student **- LMS**
Task: **Julie wants (Advertisements)** / Matching topic sentences to text
Time spent completing the task: 17 minutes

General notes

1   Reads the rubric superficially.
2   Reads the Example Item.
3   Reads through the items/questions quickly, focusing on what is required by each. Does not understand the meaning of the word '*valuable*' in Item 1, but this does not prevent her from giving a correct answer to the item. Misunderstands the key word '*entertainment*' in Item 8, which, however, results in her failure to answer the item correctly. She has no problems in understanding the rest of the items.
4   Reads through the options/advertisements silently, one by one.
5   Tries to identify the topic of each advertisement, not worrying about unknown words.
6   Responds to 6 (out of 10) items correctly while reading the options for the first time (Items 2, 3, 5, 6, 7 and 9).
7   When she has read through all options, she returns to the three items for which she has not been able to find a suitable answer (Items 1, 4 and 8).
8   Responds to two of the three yet unanswered items, still leaving the answer to Item 1 open.
9   Finally, she responds to Item 1, as well.
10   She gives a correct answer to 8 (out of 10) items.
11   She gets two items, Items 8 and 10, wrong.

Notes item by item:

<u>Item 1</u>
Similarly to LS, she does not know the word '*valuable*' in the item ('*Jack wants something old and valuable*'). When reading through the options for the first time, she considers Option C, which advertises '*antique style*' furniture, as a possible answer to both this item and, as a result of misunderstanding Item 8 ('*Peter wants home entertainment*'), to Item 8. Although she does not particularly worry about unknown words in the advertisements, the unfamiliar word '*valuable*' in the item makes her uncertain to the extent that she abandons responding to the item until she has answered all the other items on the task. By that time, she has already used the other plausible option, Option C, to answer Item 8, while it did not cause any difficulty for her to eliminate the remaining options on the basis of their content and, thus, eventually get the item right. Nevertheless, in the follow-up questionnaire, she assesses the item as one of the two most difficult items on the task, rating both "5" on the 1-6 scale.

<u>Item 2, 3, 5, 6 and 7</u>
Gives a correct answer to these items very easily, recognizing in each case the lexical overlap between the item and the advertisement (*sandals – shoes, feet*; *eat out with a boyfriend – romantic meal*; *toothache – dentist; a band for a party – music; hairdo – hair salon*).

Item 9

Similarly to Items 2, 3, 5, 6 and 7, she answers the item recognizing the lexical overlap between the item and the correct answer (*'glasses'* in the item, *'eyewear'* in the heading of the ad). However, unlike in the case of the above five items, she does not recognize the relationship between the item and the suitable option as soon as she has read the advertisement. As her report shows, she does not understand the information in / the meaning of the advertisement proper, including one of the key words *'spectacles'*, and can only respond to the item when she recognizes the word *'eyewear'* in the heading of the ad.

Item 4

She has difficulties understanding the meaning of the advertisement. However, she identifies key words in it (most importantly, the word *'Turkey'*, but also *'6 persons'*), which enables her to get the item right. However, she answers the item only after reading through all options once and answering most other items on the task.

Item 8

She has difficulties understanding the meaning of the item (*'Peter wants home entertainment'*), because of her problem with the word *'entertainment'*. As a result, the item is among those three where she abandons response until she has already answered most other items. Eventually, she selects a wrong answer to the item. One of the main reasons for this might be that, as her report shows, she interprets the word *'entertainment'* in the item as *'equipment'* (which word, in fact, appears in the correct answer). Then she tries to find a suitable option accordingly, that is, one that advertises 'home equipment' instead of 'home entertainment'. Thinking in this way, she associates 'home equipment' with *'furniture'*, advertised in one of the options used as a distractor, specifically, Option C, which leads to her selection of an incorrect answer. She assesses the item, along with Item 1, as the most difficult item on the task.

Item 10

She gets the item wrong. She selects her answer to the item while she is reading through the options for the first time, and she does not think of changing her answer later, either, which means she is fairly confident about its correctness. As her report suggests, the source of her wrong choice of answer might be twofold. First, her understanding of the meaning of the item (*'Roger wants pictures for his business'*) shows some uncertainty. She is unsure about the meaning of the preposition *'for'*. Second, she clearly associates the word *'picture'*, used in the item, with 'drawing' (rather than 'photography'), and then 'drawing' with the word *'pencil'*, which is used in the heading of an incorrect option, specifically, Option J (*'PENCIL PORTRAITS'*). As a result, she identifies, without any apparent hesitation, the topic of Option J as a *'kind of drawing studio'*, and selects, wrongly, the option as one that fits this item. It is also clear from her report that she does not even consider the possibility of choosing, as a suitable answer to this item, an advertisement in which the word 'photography' is used (like, for example, Option M, the correct answer). This, at the same time, shows that she does not necessarily try to identify matches between an item and the answer to it on the basis of overlapping vocabulary. Her way of approaching and thinking about the item as described above is, apart from her report, further supported by the fact that, in the follow-up questionnaire, she assesses the item as one of the 5 easiest items on the task, rating it "1" on a 1-6 scale.

Transcript No 4
Protocol produced by a low/middle level student - **LMS**
Task: **Caught out in the rain** / Matching headings to text

(Reads the instructions to the task.) I have to read the text / there are titles for it / and that has to be matched . / there is an extra title . / and there is an example.

(Skims through the text for 54 seconds.) Yes / I look at this, too.

R: Could you speak up a little?

Now I read through the titles / what titles there are.

(Reads through the options / paragraph headings for 28 seconds and then checks the Example heading against Paragraph 0.)

For 1 [Paragraph 0] . that's / that's indeed about the problem . / so that's / perhaps I would also write that there . / [Item] 1 /

(Reads Paragraph 1 and tries to identify a heading for it for 48 seconds.)

R: Remember to keep saying aloud what you're thinking about.

Well / now I'm thinking about . which one I could . perhaps put . for [Item] 1 . / but let's go on / I don't know / well I don't understand exact / I only understand small snatches from it and then on the basis of that . I don't know yet / so first I'll read through the whole / I think /

(Reads silently for 3 minutes 45 seconds.) Ugh! Well / this is difficult . . / Well / then (Turns to the Options.) . . . . I'm trying to understand what they could mean . but . . escaping / '*escape*' . . . .

R: Remember to keep talking.

Well / the trouble is I don't understand the text . . / Then let's see / (Re-reads the Options.) In [Option] B . / some tight escape / well it's about something to do with escape / [Option] C . . . . that's two I don't know what / that . / that the . the common people too / that is / that the . / they use . . the office buildings . . / or something like this / . . . [Option] D is the . the best way . . / escape / shelter . / to find shelter from . or in the rain / . . . . [Option] E / The man's . house is his . castle / . . . [Option] F / An accidental something / I don't know what / . . . [Option] G / Possible . short and cut . / I have no idea . . / There was this kind of . / the . / the man's house is his castle . / and . / in one of them it was mentioned that . / what his house looked like and how . . / somewhere here near the end / (Checks Option E, '*An Englishman's home is his castle*', against the text.) (Pause – 44 seconds) Perhaps it's [Item/Paragraph] 5 / (Pause – 17 seconds)

R: What are you thinking about now?

That in [Option] E it was that his house is his castle and for that perhaps [Item] 5 is suitable because this is mentioned in that . / that the hou / well that's the one in which his house . / the house is talked about / (Pause – 34 seconds)

R: Remember to keep talking.

I'm just looking at what I understand and what I don't understand from it. At [Item] 1 (Pause – 42 seconds) / well / that talks about something sudden and . accidental and that perhaps . matches [Option] F. (Reads Paragraph 2 for 19 seconds.) Some wo / middle-aged woman . . tries to help (Pause – 25 seconds) / who knew exactly what I wanted because I was already the fifteenth person who . . used . her something on this morning. (Pause – 55 seconds)

R: Remember to keep saying aloud what you think.

I'm just thinking about [Item] 4 . / because perhaps that's the escape . / the escape-related thing . . . . / because the woman didn't pay attention to him and that . / that . . . . / and that he went out to the street in the Manchester rain . / perhaps he escaped there but it's not for sure /

(Re-reads and thinks about Paragraph 6 for 43 seconds.) Well . / I'm here in England / Britain . and . . this is only / .?. / '*land*' is something . / to land / '*get off*' / to get out / to get off . / No, I have no idea (Pause – 16 seconds) Well . . . . / I look at [Item] 2.

(Re-reads Paragraph 2 and tries to identify more matches between headings and paragraphs.) (Pause – 76 seconds)

R: Remember to keep saying aloud what you think.

Well / [Items] 4 and 5 are already done . / I mean I've made guesses about them and that . . . . / I don't think / I have no idea (Pause – 35 seconds)

R: What are you thinking about now?

Well / I'm trying . [Item] 2 because that's the one I understand best but / (Pause – 16 seconds) but for that I can't find a title so . / I don't know . / I give it up /

R: So you would finish it here at a real exam?

Yes, I think so.

Low/Middle Level Student **- LMS**
Task: **Caught out in the rain** / Matching headings to text
Time spent completing the task: 22 minutes

General notes

1 Reads the instructions to the task.
2 Skims through the text before reading the options/paragraph headings.
3 Checks the Example heading against Paragraph 0.
4 Tries to respond to the items in the order they are presented on the page from 1 to 6.
5 After an unsuccessful attempt to respond to Item 1, she reads the whole text more carefully.
6 Has difficulties in understanding the main ideas in most paragraphs of the text.
7 Has apparent vocabulary problems, resulting in her failure to understand the necessary information for a correct answer, in the case of Options C, F, and G.
8 Gives a correct answer to one item, Item 4, gets Items 1 and 5 wrong, and fails to respond to Items 2, 3 and 6.

Notes item by item

Item 1
She gets the item wrong.
The main reason for this is that, as her report clearly shows, she fails to understand the main idea in the paragraph. Secondly, she does not understand the meaning of the correct answer to the item (Option G, '*Possible short-cut*'), as she is not familiar with the key phrase '*short-cut*' used in it. However, she identifies the lexical overlap between the paragraph and an incorrect option (the adverb '*suddenly*', used as the second word in the first sentence of the paragraph, appears in the form of an adjective in Option F, '*A sudden obstacle*'), which makes her find the incorrect option a suitable heading for the paragraph. (As she says, "well, that [Paragraph 1] talks about something sudden and accidental and that perhaps matches [Option] F.")

Item 2
She fails to respond to the item.
She only partially understands the main idea in paragraph 2, which may result from her failure to understand important information in the preceding (two) paragraph(s). Besides, she does not understand the meaning of the correct answer (Option F, '*A sudden obstacle*'), as she is not familiar with the key word '*obstacle*' in it.

Item 3
She fails to respond to the item.
Unfortunately, with respect to this item, her report provides no specific data on the difficulties she had in understanding the meaning of either the paragraph or the correct answer (Option A, '*A trick – will it fail?*'), apart from a general comment related to her understanding the text as a whole (As she comments, "The trouble is I don't understand the text.").

Item 4

She gives a correct answer to the item.

She understands the main idea in the paragraph, as well as the necessary information in the correct answer. Although she appears to have understood the meaning of the correct answer only partially (Option B, '*An unexpected narrow escape*'), she is familiar with the key word '*escape*' in it, which enables her to identify the match between the paragraph and the correct heading.

Item 5

She gets the item wrong.

She does not understand (or misunderstands) the main idea in the paragraph, which is the main reason for her getting the item wrong. Besides, however, she does not fully understand the meaning of the correct answer, either (Option C, '*Two approaches to public use of office buildings*'), as she is unfamiliar with the key word '*approaches*' in it, which may contribute to her failure to answer the item correctly.

Item 6

She fails to respond to the item.

As is clear from her report, she does not understand the main idea, included in the key sentence '*Get off my land*', in the paragraph. Besides, her understanding of the meaning of the correct answer (Option E, '*An Englishman's home is his castle*') is also only partial. The reason for the latter problem seems to be her superficial reading, her overlooking details in the option.

In the follow-up questionnaire, she assesses the task as "very difficult", rating it "6" on the 1-6 scale. As for the items, she assesses those items which she has not been able to respond to, that is, Items 2, 3 and 6, as "very difficult", rating each "6". She assesses Item 5, which was the first item she answered, as the easiest of the 6 items on the task, rating it "4".

Q-matrix A

The relationship between the items and the variables
As determined by Content Analysis

Item	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22
1	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
5	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0
8	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0
9	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0
11	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
12	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
13	0	0	0	0	0	0	1	0	0	1	1	0	0	1	0	1	0	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
15	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
16	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0
17	1	0	1	0	0	0	1	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0
18	1	0	1	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0
19	0	0	0	1	0	0	1	0	0	1	0	0	0	0	1	0	0	0	0	1	0	0
20	0	0	1	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0
21	0	0	1	1	1	0	1	0	0	0	0	1	0	1	0	1	0	0	0	0	0	0
22	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	0	1	0	0	0	1	0
23	1	0	0	0	0	0	1	0	1	0	1	1	0	1	0	0	1	0	0	0	1	0
24	1	1	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	1	0	0	1	0
25	1	0	1	1	0	0	1	0	0	0	1	1	0	0	1	0	0	0	0	0	1	0
26	1	1	0	0	0	0	1	0	0	0	1	0	0	0	1	0	0	1	0	0	0	0
27	1	1	0	0	0	0	1	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
29	1	1	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	1	0	1	0
30	1	1	0	1	0	0	1	0	1	0	0	0	0	0	1	0	0	0	0	0	1	0
31	1	0	0	0	0	0	1	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0
32	0	0	1	0	0	0	1	0	1	0	0	0	0	0	1	0	1	0	0	0	1	0
33	1	1	1	1	1	0	1	0	0	1	0	0	0	1	0	1	0	0	0	0	0	0
34	0	0	0	1	1	0	1	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0
35	0	1	0	0	1	0	0	1	0	1	1	0	0	0	1	0	0	0	0	1	0	1
36	0	1	0	0	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	1	0	1
37	0	1	0	1	1	0	0	1	1	0	0	0	0	1	0	0	0	0	1	0	0	0
38	0	1	0	0	1	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0	1
39	0	1	0	0	1	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0	1
40	0	0	0	1	1	0	0	1	1	0	0	0	0	0	1	0	0	0	0	0	1	0
41	0	0	0	0	1	0	0	1	0	0	0	1	0	0	1	0	0	0	0	0	1	0
42	1	1	0	0	1	0	0	1	1	0	0	0	0	1	0	0	1	0	0	0	0	1
n	13	12	7	10	12	10	24	8	9	13	6	6	10	17	15	16	8	4	4	3	11	5

Q-matrix B

The relationship between the items and the variables
As determined by VPA

Item	v1	v2	v3	v4	v5	v6	v7	v8	v9	v10	v11	v12	v13	v14	v15	v16	v17	v18	v19	v20	v21	v22
																			Variables			
1	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0
2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
3	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
4	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
7	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	1	0	0	0	1	0	0	0	0	0	1	0	0	0	0	0	0
10	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	1	0	0	0	0	0
11	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0
12	0	0	0	0	0	0	1	0	0	0	0	0	0	1	1	1	0	0	0	1	0	0
13	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0
15	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0	1	0	0	0
18	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0
20	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0
21	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0	0
22	0	0	0	0	0	0	1	0	0	1	0	1	0	0	1	0	0	0	0	1	0	0
23	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
24	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	1	0	0
25	0	0	0	0	0	0	1	0	0	0	0	1	0	0	1	0	0	0	0	0	1	0
26	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	1	0
27	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	1	0	0	0	0	0	0
28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0
30	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
31	0	0	0	0	0	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
34	0	0	0	0	0	0	1	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0
35	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	1	0	1
36	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	1
37	0	0	0	1	0	0	0	0	1	1	0	0	0	0	1	0	0	0	0	0	0	0
38	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	1
39	0	0	0	0	0	0	0	1	0	1	0	0	0	0	1	0	0	0	0	0	0	1
40	0	0	0	1	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	1	0	1
41	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0
42	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
n	0	0	0	2	0	9	21	4	4	18	0	8	0	11	15	14	2	0	2	9	3	5