Technische Universität Chemnitz
Philosophische Fakultät
Englische Sprachwissenschaft
Prof. Dr. Josef Schmied

Magisterarbeit

Testing Techniques for English Tenses and Prepositions Compared:

A Computer-Based and a Paper-and-Pencil Test of Grammar

Vorgelegt von Katrin Uhlig
Matrikelnummer: 23364
14. 07. 2005

0. CONTENTS

Acknowledgements

I would like to thank my family and friends for their support and patience during the timeI was working on this paper. Furthermore, I want to thank Petra Naumann and John McCulloch for helping me with practical and theoretical questions concerning English language testing in general and placement testing at Chemnitz University in particular. Special thanks go to Kerstin May and Katrin Voigt for the many helpful discussions an the time they spent proof-reading this paper.

1. Introduction

The crucial role language testing has to play within the field of Applied Linguistics (AL) has been acknowledged by a number of researchers. Davies refers to language testing as a means of compelling "attention to the meaning of ideas in linguistics and applied linguistics" (1990: 2). According to him

> What language testing offers applied linguistics is:
> 1 an operationalizing of its theoretical constructs;
> 2 a means of establishing goals and standards for teaching courses and syllabuses;
> 3 a methodology for carrying out empirical research in applied linguistics, whether that research is language testing (*tout court*) research; investigations in language acquisition, judgements, intelligibility studies, comprehension and use; or comparative experiments in language teaching methodologies and materials
> (ibid.).

Like Davies, Bachman acknowledges the crucial role of language testing in supplying vital information for the field of (applied) linguistic research when he mentions one important use of language tests to be that of an indicator "of abilities or attributes that are of interest in research on language, language acquisition, and language teaching" (1990: 54).

The practise of testing in general can be assumed to date back to the seventeenth century; records of an assessment of people's linguistics abilities can even be found in the Bible (c.f. Spolsky 1995: 9ff.). However, it was not until the Second World War that "the idea of language testing as a distinct activity […] existed" (Baker 1989: 29). Early approaches to language testing as a "systematic 'scientific'" practise were based on the one hand on the principles of psychometric testing, i.e. the use of closed testing questions in combination with a "ready-made set of methods and criteria for analysing and evaluating language tests", and on the other hand on the ideas of structural linguistics (c.f. 29ff.).

Present-day approaches to language testing differ considerably from those of the psychometric era. Baker notes two "shifts of interest" in language teaching, which had an impact on language testing since the end of the 1970s (1989: 1). Among them is a growing interest in ESP (English for Specific Purposes) and, in the course of this development, in "the use of language to achieve specific tasks in specific situations" (2). This strong emphasis on language in use is reflected in a more general fashion in frameworks of communicative language competence or ability , which shall be discussed in closer detail in chapter 3.3 below, and of communicative language testing. According to Weir, communicative test tasks need "to be as direct as possible,

incorporating as many of the critical features of real-life language use as" can be included in such tasks (1993: 28).

Besides focusing on communicative aspects, recent discussions of the concept of language testing focus on a variety of issues such as test fairness, and its influence on test validity or the ever increasing significance of computer-based language testing (c.f. Kunnan 1999). One of the most obvious reasons for computerized tests becoming more and more frequently used in areas such as English as a Foreign Language (EFL) is the fact that such formats can help to save time and effort especially in contexts where the number of test persons is particularly high (c.f. Fulcher 1999: 289f.). As will be described more closely in chapter 2.1 computer-based tests of English have, thus, gained considerable importance in university settings, particularly with reference to placement and entry testing.

The Department of English and American Studies' Practical Language Unit at Chemnitz University of Technology administers an English language placement test (TUC-PT) every semester. This test is entirely paper-and-pencil-based and there have been continuing discussions about converting the test or parts of it into a computer-based format. As a first step into this direction, a web-based placement test (WBPT) which provides students with the opportunity to practise for the grammar components of the TUC-PT has been developed. In order to investigate the potential of the WBPT to serve as a means for future administration of the grammar sections of the TUC-PT on the computer, it is important to describe and compare both tests in a detailed way. Thus, the main emphasis of this paper, will be on the analysis and comparison of both tests with a view to a number of different aspects.

After a brief account of some background issues concerning the field of language placement testing and a discussion of the main theoretical concepts that form the basis for the comparison of the two test instruments has been provided, the two will be formally described and compared. For this purpose, a framework for the description of language tests according to five classification features by Bachman (1990) will be taken as a basis. As a next step, a detailed analysis and comparison of individual items from the two tests with reference to their grammatical contents will be carried out. This comparison will be restricted to the testing of tenses and prepositions as two major areas of English grammar and will be conducted along the following two lines. First, a text-based multiple-choice task in the WBPT will be compared with two sentence-based multiple choice sections of the TUC-PT. Besides a detailed focus on individual items from these sections, their composition, the choice of distractors, and the grammatical contents being assessed, questions concerning the importance of the amount of context

provided for multiple choice test items will be of major interest in this section. Second, a C-test task form the WBPT will be compared with a text-based multiple choice section out of the TUC-PT. In this part of the paper, the analysis of example items from both tests will serve as a basis for a comparison of two different techniques. Special emphasis will be given to the potential of the C-test as an instrument for the testing of tenses and prepositions in comparison to a traditional grammar test method such as multiple choice. For all the items discussed, the percentages of the testing population that selected or provided an incorrect solution will be given. Although this paper is not primarily concerned with an analysis of learner errors, the amount of the target group that had problems with a respective item, and especially the type of problem, will nevertheless provide interesting insights concerning the items discussed, especially with regard to the usefulness of distractors or the item itself.

In spite of a the strong focus on the testing of grammar in this paper, the TUC-PT in its entirety will be described in some detail in section 2.2 as well as within the formal comparison of the two tests in order to underscore the status of its grammar sections as just one component of a broader testing battery which focuses on a variety of aspects of communicative language competence or ability. Serving as a practise version of the grammar sections of the TUC-PT, the WBPT does not claim to allow any valid inferences about test takers' language proficiency or communicative ability. Rather it needs to be viewed as a reflection of the significance that is attached to the explicit testing of grammatical competence by the TUC-PT which can be explained by the particularities of the university subject the TUC-PT is intended to be a placement instrument for.

2. Language Testing in University Settings

In the following, aspects related to computer-based language testing at university level will be discussed and a brief overview of the individual components of the TUC-PT and the WBPT will be given, in order to provide a basis for the analysis to follow.

2.1 The Growing Importance of the Computer

Since the 1990s there has been a growing research interest in computer- and/or web-based language assessment notable. With the ever "increasing availability of personal computers, along with increased computer familiarity", the possibilities for actually administering computer - or web-based language tests "on a large scale" have become relatively widespread (Bachman 2000:9). As a result, computerized language testing has come to play a quite pivotal role in several language testing contexts. It is

not only in the field of standardized proficiency tests of English as for example the Test of English as a Foreign Language (TOEFL) that computer-based testing has been introduced, but also in a number of other contexts such as the testing of language skills for entry into higher education programs or placement testing at university level.

Comparisons of computer-based and conventional paper-and-pencil language tests suggest that both testing forms have their advantages and drawbacks. One of the obvious advantages of a computer-based tests is the fact that they can, under certain circumstances, be used to economize the testing process in a remarkable way. Given the high numbers of students enrolling at universities every year, this is probably one of the most important reasons why so many higher education institutions have recently started to administer computer-based language tests. Especially in US contexts it seems to be relatively common to administer university entrance or placement tests in English or other subjects via the computer. Shippensburg University in Pennsylvania for example conducts a computer-based English reading placement test twice every year.[1] Another, very interesting example of English language placement testing via the computer can be found at Dakota State University in Madison, South Dakota. There, computer-based placement testing is characterized by an orientation towards two different target groups. First, there exists an English reading and writing placement test for native-speakers of English. By means of error identification techniques the latter of the two tests combines assessment of textual and stylistic competence with measures of grammatical knowledge.[2] Second, Dakota State University also offers a computer-based ESL placement test which consists of a grammar/usage, a reading and a listening section. The test uses an adaptive format and "measures multiple levels of English proficiency, ranging from near-beginner to near-native speaker ability" (*Dakota State University ESL Placement Test)*. It does so mainly by using a variety of multiple choice test items.

As a matter of fact, university English language placement tests in non-English-speaking countries such as Germany in general only have an ESL orientation. A very interesting example for a computerized ESL placement test at a German university is administered by Dresden University of Technology. The placement test of the university's Department of English and American Studies consists of a writing, a grammar and a speaking section. The grammar part of the test uses a computer-based format and consists of "3 elements (grammar, vocabulary, pronunciation)" (*Dresden University of Technology - Placement Test (WS 2004)*). Apart from the placement test,

---

[1]For more detailed information about computer-based placement testing at Shippensburg University c.f. *Placement Testing at Shippensburg University*
[2]For more detailed information about computer-based placement testing at Dakota State University, c.f. *Dakota State University Placement Test: Sample Questions,* and *Dakota State University ESL Placement Test*

the Institute of English and American Studies at Dresden University of Technology also conducts an entry test for all incoming students. Like the placement test, this entry test also contains a computer-based grammar section focusing on "tenses, general grammar, vocabulary and pronunciation" (*Dresden University of Technology – Entry Test in Winter Semester 2005/2006*). It is interesting to note that the English language placement test at Dresden University of Technology is explicitly based on an established framework of levels of proficiency or difficulty. The Department of English and American Studies states the following: "the level of our entry test is oriented towards levels B2/C1 of the 'Common European Framework of Reference for Languages'" (ibid.).[3]

2.2 Language Testing at the Department of English and American Studies at Chemnitz University of Technology – The TUC English Language Placement Test

Like the Department of English and American Studies at Dresden University of Technology, the English Language and Linguistics Department at Chemnitz University of Technology (TUC) conducts an English placement test for all incoming students who intend to study subjects that are related to English Language and Linguistics. These subjects include English and American Studies as a major subject, English Linguistics, English Literature and American Literature as minor subjects or Foreign Languages in Adult Education which generally combines English and another European language such as French or Italian. The placement test is administered twice every year by the Department's practical language unit. There exist two similar versions of the test, one of them for the winter the other for the summer semester. For the description and analysis provided by the present paper only the version of the placement test that is administered in the winter semester will be taken as a basis. It consists of three sections: The first section is a writing activity and requires the test takers to write a letter on a general topic. The second part consists of a paper-and-pencil test which combines a spelling section with three tasks focusing on a variety of grammatical issues, this second part also contains a reading comprehension section and a component that deals with the terminology of English grammar. The spelling section requires test takers to identify and correct spelling mistakes in five sentences, without any reference to the number of mistakes contained in the sentences being made. All of the grammar and grammar terminology tasks employ varieties of the multiple choice format. The reading section consists of a news text in which five sentences are missing. These sentences are

---

[3]For more information about the Common European Framework of Reference for Languages, c.f. *ALTE framework and can-do*.

provided at the bottom of the text and it is the test takers' task to fill them in at the appropriate place in the text. Furthermore there are several statements concerning the text given and test takers have to decide whether these statements are true or false. Finally, the reading section of the placement test contains a multiple choice task dealing with lexical items from the text for which testees have to choose a suitable paraphrase from a list. The third part of the placement test is an oral interview which is administered after the first and second part and largely serves as a means to refine decisions taken about a test person's language ability level on the basis of his/her scores in the first two sections of the test. Contrary to the testing practise of the universities mentioned above, the written sections of the TUC-PT are entirely administered in a paper-and-pencil-based format.

The number of students who enrol at Chemnitz University of Technology in the subjects mentioned above and who consequently take the English language placement test is rising form semester to semester. Whereas the number was about 130 in the summer semester 2004, 140 students took the test in the winter semester 2004/2005. As has been emphasized by Bachman, the use of computers in language testing can be a useful means to economize testing and scoring procedures (2000: 9). This might be especially helpful in settings where the high number of test takers would otherwise demand a great scoring effort. As has been outlined above, the number of candidates taking the placement test administered by the Department of English and American Study's practical language unit every semester is steadily rising. Accordingly, questions concerning the potential of the TUC-PT to be administered via the computer to increase the effectiveness of the scoring process are very interesting ones.

2.3. Computer-Based Language Testing at Chemnitz University of Technology - A Web-Based Placement Test of English Grammar

Like Bachman (2000: 9), Bugbee refers to the ever increasing importance of computer-based testing and gives a summary of the advantages of this type of assessment. According to him,

> The advantages of administering tests by computer are well known and well established. Computerized tests have been shown to reduce testing time (Bunderson, Inouye, & Olsen, 1989; English, Reckase, & Patience , 1977; Green, 1988; Olsen, Maynes , Slawson & Ho, 1986; Wise & Plake, 1989); to obtain more information about test takers (Wise & Plake, 1989); to increase test security (Grist, Rudner, & Wise, 1989); to provide instant scoring (Bugbee, 1992; Bugbee & Bernt, 1990; Kyllonen, 1991; Mazzeo & Harvey, 1988); and to be scheduled more easily than paper-and-pencil-administered tests (Bugbee, 1992; Hambleton, Zaal, & Pieters, 1991; Wise & Plake, 1989). (1996: 282f.)

Not only does this provide a quite concise overview over the positive effects of computers on the language testing process, a closer look at the sources that Bugbee uses to underpin his statements reveals at least two other interesting issues. Firstly, the publication dates of the studies and articles cited by Bugbee clearly indicate that although computer-based language testing is quite frequently described as a phenomenon of the 1990s (c.f. e.g. Alderson & Banerjee 2001; Bachman 2000) its roots can be assumed to date back much further. Furthermore, the topics, the articles and studies deal with, show that the field of language testing can be and has been approached from a variety of angles. These include educational as well as psychological issues.[4] The facts mentioned by Bugbee and a number of other researchers seem to indicate that computerized formats can have a high number of advantages with regard to testing procedures in general and language testing in particular. Therefore, for those sections of the TUC-PT that focus explicitly on English grammar and its terminology, a web-based practise version has been developed. This web-based practise placement test will in the following be referred to as web-based placement test (WBPT) only, keeping in mind that it lacks a number of criteria that are typical for a placement test in the strict sense of the word. These will be further discussed in chapter 4.1.

In the following, a brief account of the composition of the WBPT will be given. The test consists of five parts. For each of the five test sections there exist several versions. For each of the sections, test persons are supplied with one of these versions via a random principle. The part of the WBPT consists of a text-based multiple choice gap filling activity. The text contains 25 gaps and for each gap there are 3 alternatives given. The second part of the WBPT is a C-test based on a relatively short text with 25 gaps the test takers have to fill by typing in the missing letters of halfwords. The third section of the WBPT is a multiple choice activity which requires the subjects to identify erroneous structures in English dialogues. Every test person is presented with two dialogues. These dialogues are subdivided into five parts and each of these parts contains four underlined words or phrases, with one of them being incorrect. In this task the test person has to identify the incorrect word or phrase for each part of the dialogues. Like section three and section one of the WBPT the fourth part also uses a multiple choice format, however it does not directly deal with test-takers' knowledge of the English language as such but with their capability to handle metalanguage, in this case the terminology of English grammar. The tasks in section four are based on British and American English letters. The structure of this part of the test is in many respects similar to section three just mentioned. Each test person is presented with two letters

---

[4]For a more detailed overview of the topics dealt with in the articles and studies cited, please refer to the bibliography provided in Bugbee (1996).

which are subdivided into 5 parts. In most of these parts, except for one case, in which the test taker has to identify a sentence subject, there is one word or phrase underlined. Below each of the five paragraphs of the two letters, test persons are presented with four statements about the underlined words or phrases. These statements deal with English grammar terminology. One of these statements is true, the other three are false. It is the test person's task to identify the correct statement. The fifth and final part of the WBPT is again a multiple choice task. It consists of five short dialogues dealing with every day conversation topics, such as shopping, holidays, talking to old friends etc. Each dialogue contains five gaps which deal with a particular grammar point, for each of the gaps there are four alternatives given. In contrast to section one of the WBPT, in this part the alternatives are not words or phrases but whole sentences.

Test persons are allowed 60 minutes time to do the test. After successful completion test takers receive a summary of their correct and incorrect answers and a percentage score. The WBPT does not use categories of success or failure when presenting final scores. Rather the WBPT provides test takers with a certain estimation of their proficiency level in English grammar. Nevertheless, it uses a fixed score value of 60% to differentiate those test persons with fairly high levels of grammar competence from those with fairly poor knowledge of the subject. This latter group is advised more grammar practise as further preparation for the TUC-PT.

It needs to be maintained here that the WBPT is a web-based test instrument which can be used by anybody at any time under any circumstances, given the technological means needed for doing so are at hand. Thus, one needs to keep in mind that for any of the test takers, there exists the potential opportunity to use certain aids, such as (on-line) grammars, dictionaries etc. Although the time limit provided to complete the test will probably not allow a test taker to look up every unknown word or to check all of the alternatives in the multiple choice sections in a grammar book the testing situation provided by the WBPT cannot be compared with the situation typical for tests which are used to make major decisions about an individuals language knowledge or ability.

As mentioned earlier on, the WBPT functions mainly as a practice opportunity for first semester students of English and American studies at Chemnitz University of Technology, or one of the related subjects cited above, who have to take the TUC-PT at the beginning of their studies. The main target group of the test are first semester university students who, in general, have completed their school education with the 'Abitur' or an equivalent. Because of the fact that a great number of these students were referred to the WBPT either by notes in the English and American Studies university

calendar or on the department corridor and a number of other highly frequented places on campus as well as by the tutors of the Department's English practical language program it can be supposed that a relatively high number of individuals within the testing population that was analysed for the purposes of this paper belong to this assumed target group. Most subjects in the target group were thought to be of German nationality and consequently to have German as their mother tongue or one of their mother tongues. This is also reflected in some of the items of the WBPT which focus on German-English interference problems. However, it needs to be kept in mind that every semester, there is a certain number of foreign students enrolling in English and American Studies or a related subject at Chemnitz. It is thus important to assure that the number of tasks and items focusing on special grammar problems that are typical for German learners of English do not outweigh the general grammar topics that are focused on by the test.

The fact that the WBPT is delivered via the WWW and therefore can be accessed by basically everybody all over the world results in a potential for the target group to become relatively heterogeneous. In order to have a certain control of the composition of the group of people taking the test, the WBPT requires the test takers to complete a short questionnaire focusing on a number of demographic facts as well as information about testees' language background and experiences and their educational level. Precisely, the questionnaire asks test takers about

(a) their sex

(b) their age

(c) their mother tongue

(d) their level of education.

Furthermore the test taker is asked

(e) for how long (s)he has been studying English so far

(f) whether (s)he has already taken the Chemnitz University of Technology English language placement test at an earlier point of time and if so to state the date when (s)he did so

(g) whether (s)he has ever spent a longer period of time (at least three months) in an English-speaking country and

(h) whether (s)he has ever taken any standardized English proficiency test such as the TOEFL or one of the Cambridge exams.

3. Theoretical Background

The following sections will give a theoretical overview of the concepts of

grammar testing, placement testing as well as grammatical competence and the concept of language proficiency in order to provide a solid basis for the comparison of the WBPT and the TUC-PT with regard to formal as well as linguistic characteristics of the tasks and items of the two tests.

3.1 Grammar Testing

Early approaches to grammar testing, such as the one provided by Robert Lado, were deeply rooted within structural and early contrastive linguistics (Rea-Dickins 1997: 87).[5] For Lado, parallel grammatical categories, such as "gender, number, person, case, voice, subject-object relation, head-modifier structure and subject-predicate constructions" (1961: 145) in different languages were of very high interest for the development of grammar tests. Lado underscores the high influence of the grammatical structures of a person's mother tongue on his/her learning and understanding of the grammatical patterns of a foreign language. For him, testing grammar in a foreign language seems first of all to be concerned mechanisms of transfer of a "set of habits" from a learner's L1 to the L2 (c.f. 146ff.). It is especially those instances of such transfer which do not result in correct or satisfactory utterances, i.e. instances of negative transfer that should, in Lado's view, be focused on by language tests in general and grammar tests in particular. With his approach to grammar testing, Lado combines structuralist as well as behaviourist principles; his way of describing the origins of potentially problematic patterns within the grammar of a language for a given test person is strongly in accordance with the early principles of contrastive analysis, particularly the Contrastive Analysis Hypothesis in its strong version. In its strong form this hypothesis is based on the claim that it is possible to "predict and describe the patterns that would cause difficulty in learning, and those that would not cause difficulty, by systematically comparing the language and culture to be learned with the native language and culture of the student" (Lado 1957 cited in Zhang, n.d.). Relating this to grammar, it becomes possible to suggest that the above approach clearly regards a systematic comparison of the grammatical structure of a learner's L1 and that of the foreign language to be learned as one of if not the most crucial prerequisite for developing grammar tests. Such a comparison should accordingly focus on those areas of grammar that are especially prone to negative transfer and following from this a test has to be developed which focuses on exactly these problem areas – or, as Lado puts it - "testing control of the problems is testing control of the language" (1961: 24).

---

[5]For a detailed account of the history of grammar testing c.f. Rea-Dickins (1997).

Lado defines the grammatical structure of a language as "the patterns of arrangement of words in sentences and the patterns of arrangement of parts of words into words" (1961: 142). Generally speaking, the main focus in Lado's approach to grammar and grammar testing seems to lie upon the word and the sentence level. This becomes especially clear if one takes into account his division of the grammatical structure of a language into several levels. These are: "*sequences of sentences, sentence patterns, parts of sentences, modification structures, parts of speech and function words, inflection, derivation, patterns of morphemes into words*, and *morphemes*" (144). For Lado, grammatical patterns constitute an important framework which seems to be fundamental for the production of (correct) utterances in a language and therefore for communication in general. Accordingly, he views grammar or grammatical patterns as "the mold from which countless utterances can be produced" (144). It thus seems that structuralist approaches to grammar testing do, above all, aim at testing the respective grammar construct as such, i.e. more or less isolated from the skills needed by a learner in order to successfully communicate in a given real-life context.

In accordance with a general tendency within language testing and teaching as well as linguistics in general from structuralist towards pragmatic and/or communicative principles, present day approaches to the testing of grammar vary considerably from those typical for the era of Lado. One of the most influential approaches to the concept of communicative grammar was provided by Leech, who distinguished three levels of grammar: the syntactic, the semantic and the pragmatic level (1983: 152 cited in Rea-Dickins 1991: 114). Leech defined "communicative grammar as 'an approach to grammar which aims to relate these three levels of description to each other'" (ibid.). For Leech, grammar as "'the formal system of a language'" and the actual use of that language, its pragmatics, are "'complementary domains'" (1983: 4 cited in Rea-Dickins 1991: 114). Relations of the formal structures of a language to its use in real-world communication are also expressed in a number of descriptions of the concept of language proficiency which will be discussed in chapter 3.3 below. Such relations are also an integral part of what has come to be known as communicative language testing. For Rea-Dickins "the communicative testing of grammar requires a task-based approach"(1991: 124). She gives a number of examples for such task-based grammar tests, one of them being the following:

> **Situation:** You have invited your colleague Maria to dinner. She hasn't been to your house before so you go to her office to tell her where you live. She isn't there so you decide to leave her a message.
> **Instruction/Task:** Write a short note to Maria giving her a clear set of instructions so that she can get form work to your house for dinner (125).

Clearly, this test-task varies considerably from more typical forms of grammar tests, e.g. the ones used in the WBPT or the TUC-PT. It might even be argued that the term grammar test is not a sufficient label for this type of task as it focuses not only on knowledge of the grammatical structures of a language and/or the ability to implement this knowledge in a certain communicative situation but also on a number of other competencies. In order to successfully complete the task, the test taker has to combine a variety of skills such as writing skills, text structuring techniques, and (s)he also has to take into account pragmatic issues such as writing for an addressee and giving information in such a way that it can easily be followed by another person. Summarizing, types of grammar test tasks such as the one just mentioned focus on an implicit role of grammar within a broader communicative task as one of the means to ensure successful communication and can be said to reflect Leech's claim of a combination of knowledge of grammatical structures and rules and the ability to use a respective language. Such implicit testing of grammar has since gained more and more importance.

In general, views of grammar as a means to achieve a particular communicative goal can be described to be consistent with the principles of communicative language testing as defined by Bachman (c.f. e.g. 1990). According to him, it is especially important for communicative language tests to allow valid inferences about a test person's real-world language ability. It is thus highly relevant in order for a test to be valid "(a) that the language abilities measured [...] correspond in specifiable ways to the language abilities involved in nontest language use, and (b) that the characteristics of the test tasks correspond to the features of a target language use context." (1991: 681 cited in Miyata-Boddy, & Langham 2000: 81). Applying this to grammar testing, the above view suggests that it is not grammar as a system of elements and patterns which should be measured by a language test but the ability of a testee to use grammar in order to achieve a particular communicative goal, preferably in the real world. Such tests do therefore not explicitly measure a person's actual grammar competence but they rather focus on the communicative skills this person possesses. There exist a number of views concerning the relation between grammatical ability and the ability to communicate in a foreign language and many researchers, among them Bachman and Palmer (1996), and Bachman (1990), clearly see grammatical competence as one component of communicative competence as a whole.[6]

Apart from the growing significance that is being attached to implicit measures of grammar, the explicit testing of grammatical knowledge or competence, i.e. testing

---

[6]c.f. Rea-Dickins 1997: 89 on the relation between grammatical and communicative competence.

grammar more or less separately from other language skills or tasks, has been an integral part of language testing for decades and this seems to continue into present-day language testing practice. Rea-Dickins points to this fact when she refers to the continuing significance of explicit grammar testing in classroom contexts (2001:22). In contrast to that, she points out that for language proficiency examinations, e.g. those "used for university admissions purposes", a shift in the significance of grammar has taken place (22f.). In the past, most proficiency examinations, e.g. the English Proficiency Test Battery (EPTB), the English Language Battery Assessment (ELBA), the Comprehensive English Language Test (CELT), or the TOEFL "included a grammar component" separate from the other components and clearly identifiable as directly focusing on the assessment of grammatical knowledge (ibid.). It is interesting that, although "some of these examinations are [nowadays, K.U.] more or less defunct" (ibid.) and "the direct testing of grammar has largely fallen out of favour"[7] (Alderson & Banerjee 2002: 89), the TOEFL as one of the most influential English language proficiency test batteries world wide continues to employ a "multiple choice subtest of structure and written expression", i.e. an explicit and separate grammar (and usage) component and it does so even in its computer-based version the TOEFL-WBT (Rea-Dickins 2001: 23). The same is true for most of the Cambridge Exams, especially those that focus on higher levels of language proficiency.

In the course of a growing research and teaching interest in the use of language as a communicative tool, which resulted in changing opinions about the nature of language learning and language proficiency, there has indeed a decrease in the significance attached to explicit and separate grammar testing taken place. However, the fact that the TOEFL and many other language test formats still use explicit measures of grammar indicates that there is still a relatively high significance attached to this field of language testing. Although the number of recent research papers and projects focusing on grammar testing is fairly low, there are some crucial questions concerning this topic which are discussed quite controversially. Among them is the central question whether grammar "should [...] be tested separately form other skills", i.e. explicitly, at all (Rea-Dickins 2001: 22).

In spite of the shift in defining the grammar construct just outlined and the resulting significance that is being attached to implicit measures of grammar, there are still a fairly high number of language learning and testing contexts where the explicit

---

[7]When talking about *direct testing* of grammar, Alderson and Banerjee focus on measures of grammar which explicitly focus on the knowledge of grammatical structure of a language. However, the terms *direct* and *indirect measurement* can also be used in a different sense. For a discussion of these meanings of the terms *direct* and *indirect measurement* c.f. n8.

testing of grammar has not lost its influence and significance. This can for a example be seen in higher education settings: Many universities in Germany as well as the USA or Great Britain still employ an explicit grammar component in their placement or admissions tests. Furthermore, the shift from structuralist towards communicative principles in defining the construct of grammar is not only reflected in implicit but also, though obviously to a far lesser extent, in explicit measures of grammatical ability in a foreign language. Rea-Dickins notes two changes in "the representation of grammar" (2001: 28) within explicit grammar tests which clearly underscore this fact: The first of these changes is characterised by a broader definition of the concept of grammar which goes well beyond sentence level and has come to include factors like "textual competence such as cohesion, rhetorical organisation" and questions of "accuracy and appropriacy of language for" specific situations and tasks (ibid.). A second change observed by Rea-Dickins consists in the fact of testing materials used for measuring grammatical knowledge as well as explicit grammar test tasks becoming more and more varied (ibid.). She provides a list of test types typically used for the testing of grammar, these include " modified cloze, gap-filling passages, matching, unscrambling words and sentences, guided short answer and summary tasks" (ibid.). Some of these task types can nowadays rather frequently be found in explicit and separate grammar tests. It can be assumed that most of the testing techniques mentioned above are more or less integrative in nature. Integrative tests require the test taker to combine various skills or aspects of the components of language proficiency when answering test items. In contrast to that, in discrete-point tests each item focuses on a single element of grammar, a single skill etc. only. Oller gives a very concise summary of what constitutes the ideas of discrete-point and integrative language testing:

> If discrete items take language skill apart, integrative tests put it back together. Whereas discrete items attempt to test knowledge of language one bit at a time, integrative tests attempt to assess a learner's capacity to use many bits all at the same time, and possibly while exercising several presumed components of a grammatical system, and perhaps more than one of the traditionally recognized skills or aspects of skills. (1979: 37).

It needs to be maintained at this point that it would indeed be an exaggeration to suggest that the use of methods such as the ones mentioned by Rea-Dickins for the testing of grammar would measure language abilities that are typical for authentic language use or that such procedures would automatically present the test taker with a language use context that mirrors a non-test situation. However, the use of such rather integrative measures for the testing of grammar shares one very important characteristic of real-life communication, namely the need for an integration of several skills in order to successfully perform a given communicative task. Furthermore, the growing degree of

integrativeness that can nowadays be observed within many tests of grammar implies the following: The distinction between explicit and implicit measures of grammar is not in all cases a clear-cut one, i.e. it is not a question of grammar being tested in total isolation from other skills as opposed to its absolutely indirect assessment as an integrated part of a broader language-related task. Rather, many present day explicit measures of grammar differ from those typical for the structuralist era in that they do not only focus on knowledge of rules. Rather, many such tasks try to incorporate a certain amount of aspects of actual language use. Moreover, the question of whether to use explicit measures of grammar or whether to test it implicitly, is in any case dependent on the particular objectives of this test. As Rea-Dickins points out, "it may [under certain circumstances; K.U.] be highly relevant to have explicit system-focused testing" (2001: 29). Such circumstances can include "for example [...] the earlier stages of language learning", situations "where course-related decisions are required" or the testing of "future teachers of English" (ibid.). It is in such situations that a grammar test might have to pay particular "attention to form and structure, knowledge of rules, or accuracy of language use" in order "to uncover whether learners have the right packaging at their disposal [...] to convey messages appropriately" (ibid.). Another field where "explicit system focused" (ibid.) grammar testing might be of particular relevance is the domain of English in university settings. Although Rea-Dickins observes a tendency towards a "more indirect[8] approach to assessing grammatical ability" in English for Academic Purposes (EAP) (ibid.), one needs to draw a clear distinction between those academic subjects for which English is a mere means of conveying technical or subject-related information and those which have the English language as their actual subject or parts of it. This is first of all the case for future English teachers as well as students of English and American Studies in general and English Linguistics in particular. These groups of students do not only need to have an extremely good command of the English language in order to successfully communicate in university and non-university settings, they also need to be able to explain, analyse

---

[8]When talking about *indirect measurement* of grammar Rea-Dickins refers to grammar being tested more or less implicitly, e.g. within a certain type of skills-based tasks as opposed to explicit grammar testing with a direct focus on grammar competence as such. (c.f. 2001: 29). The terms *direct* and *indirect measurement* are frequently used in the literature in another sense as well: "By eliciting a **performance** approximating **authentic** language behaviour" direct types of measurement "measure **ability** underlying" (Davies et 1999: 47, my underlining). In contrast to that, indirect tests do "not require the **test taker** to perform **tasks** that directly reflect the kind of language use that is the target of **assessment**; rather, an inference is made from performance on more artificial tasks" (81). Interestingly, Rea-Dickins uses the term *indirect* in a sense that seems to be in clear opposition with the definition provided by Davies et al. In order to avoid confusion, this paper uses the terms *explicit* measurement when referring to those tests which put a direct focus on grammar knowledge or competence as such and *implicit* measurement in connection with those measuring grammatical competence in a rather indirect way, e.g. in a form of task-based assessment.

and critically discuss the language in general as well as its grammatical structures in particular. It is for these reasons that it is crucial to carefully consider the goal and purpose of a language test as well as the context in which it is used in order to decide whether to use explicit or implicit measures of grammar, or a mixture of both.

As has been mentioned earlier on, the explicit testing of grammatical knowledge continues to play a quite important role in foreign language placement testing at university level. The English Language Placement Test at Chemnitz University of Technology (TUC-PT), for instance, contains three sections which very clearly and explicitly focus on grammatical ability in the sense of knowledge of the grammatical structures of the English language, the rules behind these structures and even the terminology of English grammar. Like the TUC-PT, the WBPT can be described as a test instrument which focuses on the grammatical structure of the English language in a rather explicit way. The use of a relatively high amount of explicit measures of grammar can without doubt be described as a reason of the test purpose and a result of the target group of the test, namely a group of students for who the grammar of the English language is not only a tool needed for successful communication but also a potential object of analysis and discussion. that

Before it is possible to describe and compare the way grammar is tested in the WBPT and the TUC-PT, it is of crucial importance to have a solid framework for the description and comparison of language tests at hand. For this reason, the following chapter will focus in closer detail on the concept of placement testing, its characteristics, limitations and possible overlaps with other types of language tests or language testing concepts.

3.2 Placement Testing

The term *placement* or *placement test* is very frequently associated with the intended use or purpose of a language test (c.f. eg. Fulcher 1997; Alderson, Clapham, & Wall 1995; Bachman 1990). However there is considerable overlap in defining different types of language tests, which might be illustrated by the following example: Although a number of language testing researchers describe placement testing as a type of test use or purpose, some do not do so. Among them is Davies, for whom the five main uses of language tests are "achievement, proficiency, aptitude, diagnosis and pre-achievement" (1990: 6). Interestingly, Davies utilizes the term *test purpose*, which is employed by Alderson et al. (e.g. 1995: 11) and others to refer to the use of a language test, with reference to the dichotomy norm-referencing versus criterion-referencing of language tests (Davies: 1990:6). This dichotomy is in turn subsumed by Bachman under the

headline of frame of reference (1990: 72ff.) From this it follows that there seem to exist quite a number of more or less substantial inconsistencies and overlaps within the field of language testing terminology. A more detailed discussion of these overlaps and their possible reasons and results would exceed the scope of this paper, however the above example clearly implies that for describing and analysing any given type of language test it is crucial to decide on one terminological framework which one employs throughout. For the present paper this framework will be taken from Bachman (1990) who provides a very concise and detailed scheme for test type classification, description and analysis.

Bachman classifies types of language tests according to the following "five distinctive features":

(1) intended test use or purpose
(2) test content
(3) frame of reference
(4) scoring procedure
(5) testing technique or method (1990: 70)

For Bachman, the purpose or intended use of a language test is "the single most important consideration in both the development of language tests and the interpretation of their results" (54). The classification category of test content is not so much concerned with the actual (language) content of a test but with the principles the test is based upon. These content principles can either be "a theory of language proficiency or a specific domain of content, generally as provided by a course syllabus" (71).

With reference to the content orientation of placement tests, Wall, Clapham, & Alderson observe that placement tests can have either a proficiency or a syllabus orientation (1994: 322), which shall be further discussed in chapter 4.2. This clearly underscores Bachman's assumption that language tests, in this case placement tests, can either be theory- or syllabus-based. At the same time, it becomes obvious that the distinctive features of intended test use and test content are not mutually exclusive categories. Rather, they, as well as the other features provided by Bachman, need to be treated as complementary aspects of language test description and classification. Furthermore, Wall et al.'s observation hints to questions about the relationship between placement testing and the concepts of syllabus on the one and proficiency orientation on the other hand. As shall be explained in closer detail later on, the tests that are to be analysed in this paper are proficiency rather than a syllabus-based forms of assessment. It is for this reason that a more detailed discussion of the concept of language proficiency is most interesting for the purposes of the present analysis. Thus, before a detailed description of the TUC-PT and the WBPT on the basis of Bachman's five

classification features will be given, the notion of language proficiency with special reference to grammatical competence shall be focused on briefly in the following section.

3.3 Grammatical Competence and the Concept of Language Proficiency

Defining the concept of language proficiency, which has been labelled by Huhta to be a "vague and problematic" one (1996: 218), is in fact a rather complicated undertaking. This chapter will focus on the development of different approaches to the concept since the 1960s and give a brief overview of some of the theories of language proficiency that have been most influential in language teaching and testing ever since.

Early approaches to the concept mainly focused on a "division of proficiency into reading, writing, speaking, and listening skills" (Huhta 1996: 219). Proficiency frameworks such as the ones provided by Lado (1961) mainly focus on these four skills and distinguish them "from components of knowledge (grammar, vocabulary, phonology/graphology)", but – as Bachman points out – fail to indicate "how skills and knowledge are related." (82).

Definitions and classifications of the concept of language proficiency have since changed to a considerable extent: Nowadays, there exist three basic "uses of the term **proficiency**" (Davies et al. 1999: 153). The first of them very generally refers to the concept as a "type of **knowledge** of or **competence** in the use of a language, regardless of how, where or under what circumstances it has been acquired" (ibid.). The second use of the term, puts particular emphasis on a certain aspect or field of the language or its use. In this sense, language proficiency is viewed as the "**ability** to do something specific in the language, for example proficiency in English to study in higher education [...]". The third of the uses of the term cited by Davies et al. bears an explicit reference to language testing. It looks at language proficiency as the "**performance** [...] measured by a particular testing procedure" (ibid.).

Chalhoub-Deville gives a very detailed overview of different research approaches to the concept of language proficiency. She distinguishes componential proficiency models from proficiency levels. According to her, "componential models [...] describe the components of proficiency", whereas "levels of proficiency [...] focus on depicting the progression of learners' proficiency at various stages" (1997: 4). As grammatical competence, which plays a major role for the present analysis, is nowadays frequently considered to be one component of general language proficiency (c.f. e.g. Bachman 1990: 87f.), the discussion in this chapter will be restricted to componential

models of language proficiency[9]. With reference to componential models she reviews some of the most influential conceptualisations of language proficiency, among them John W. Oller's unitary competence hypothesis (UHC) which viewed language proficiency "in terms of one general factor" (Chalhoub-Deville 1997: 4). Another term within this context is the notion of general language proficiency which is often used synonymously with UHC (c.f. Davies et al. 1999: 153)[10]. In contrast to Oller who considered proficiency in a language to be "a unitary phenomenon" (Huhta 1996: 219), later approaches to the concept ascribe a certain multi-facetedness to it: There exists for example a "considerable overlap between the notion of language proficiency and the term **communicative competence**" (Davies et al. 1999: 153).

Canale and Swain introduced their theory of communicative competence in the early 1980s. Explaining their "proposed theoretical framework" they give a concise account of the characteristics which can, according to them, be assumed typical for human communication and communicative competence as such (1980: 28f.)[11]. For Canale and Swain communicative competence consists of several aspects, these are grammatical and strategic as well as sociolinguistic competence[12] (Chalhoub-Deville 1997: 6). Canale and Swain's approach has been described by Cziko as "a 'descriptive model' whose primary concern is to define all the knowledge and skills that learners need to use the language effectively" (1984: 24 cited in ibid.).

This emphasis on language in use is also a very important characteristic of Bachman's model of communicative language ability (CLA), which can be considered as a further "componential interpretation of proficiency" (Chalhoub-Deville 1997: 7) and at the same time as an "extension of earlier models" (Miyata-Boddy; Langham 2000: 77), such as the one provided by Canale and Swain. Bachman describes CLA "as consisting of knowledge, or competence, and the capacity for implementing or executing that competence in appropriate, contextualized communicative language use" (1990: 84). Bachman's CLA model combines the knowledge of or competence in a language with strategies used to implement this knowledge in actual communication and a number of psychological and neurological processes that are part of human communication. CLA consists of the following three superordinate "components": "language competence, strategic competence, and psychophysiological mechanisms" (ibid.). Language competence in Bachman's sense consists of abilities that "can [...] be

---

[9]For a detailed account of levels of proficiency, c.f. Chalhoub-Deville (1997: 9ff.).
[10]For a more detailed discussion of Oller's approach to language proficiency, c.f. Oller (1976) or Chalhoub-Deville (1997:4f.)
[11]For a detailed overview of the characteristics of communication and communicative competence as well as the components of the latter, c.f. Canale & Swain 1980: 28-31
[12]Canale later extended the model "by distinguishing between sociolinguistic [...] and discourse competence". (Chalhoub-Deville 1997: 6)

classified into two types: organizational and pragmatic competence" (86) which will be explained in closer detail in the course of this chapter. Strategic competence is defined as "the mental capacity for implementing the components of language competence in contextualized communicative language use" (84). Finally, psychophysiological mechanisms that form the third component of CLA can be described as "the neurological and psychological processes involved in the actual execution of language as a physical phenomenon" (ibid.), such as the use of either "auditory and visual skills" for language reception or "neuromuscular skills (for example, auditory and digital)" for the production of spoken or written language (107). All the three components mentioned are then further sub-classified in a very detailed and comprehensive way.

In the following, only the components of Bachman's language competence category will be explained briefly.[13] Nevertheless, it needs to be kept in mind and acknowledged that strategic competence as well as the psychophysiological mechanisms involved in communication are integral parts of language proficiency and are in no way to be neglected. Rather, the restriction to language competence is due to the fact that the present paper puts a main emphasis on grammar testing and the explicit analysis of grammatical structures focused on by individual test items mostly on the basis of the written text. Both grammatical and textual competence form an integral part of what has been labelled language competence by Bachman. Furthermore, this paper is first of all concerned with grammar competence or knowledge which, following Bachman, clearly lies on the language competence level. He points out that "language competence comprises, essentially, a set of specific knowledge components that are utilized in communication via language" (84). He divides language competence into organizational and pragmatic competence. Pragmatic competence is then further subdivided into illocutionary competence and sociolinguistic competence. Illocutionary competence can also be described as "the knowledge of the pragmatic conventions for performing acceptable language functions" (90). In this regard, Bachman is essentially concerned with "four macro-functions: ideational, manipulative, heuristic, and imaginative" (92)[14]. Sociolinguistic competence is described by Bachman as "the sensitivity to, or control of the conventions of language use that are determined by the features of the specific language use context" (94). Such sensitivity or control can be directed towards "*differences in dialect or variety*", "*differences in register*", the natural use of a language , and "*cultural references and figures of speech*" (95-98).

---

[13]For a complete explanation of the components of CLA and their respective sub-components c.f. Bachman 1990: 81-110.
[14]For more detailed descriptions of these functions c.f. Bachman 1990: 92-94.

Organizational competence as the first of the two broad sub-categories of Bachman's language competence component "comprises those abilities involved in controlling the formal structure of language for producing or recognizing grammatically correct sentences, comprehending their propositional content, and ordering them to form texts" (87). As mentioned earlier on, Bachman divides organizational competence into grammatical and textual competence. Grammatical competence

> includes [...] a number of relatively independent competencies such as the knowledge of vocabulary, morphology, syntax, and phonology/graphology. These govern the choice of words to express specific significations, their forms, their arrangement in utterances to express propositions, and their physical realizations, either as sounds or as written symbols (ibid.).

Textual competence "includes the knowledge of the conventions for joining utterances together to form a text, which is essentially a unit of language [...] consisting of two or more utterances or sentences that are structured according to rules of cohesion and rhetorical organization" (88). An overview of the components of language competence is provided in Appendix I.

Summarizing Bachman's approach to language proficiency it needs to be maintained that probably the most remarkable characteristic of his model is the very clear focus that is being put on language in use, i.e. language competence or knowledge is not seen in isolation from the actual context within which communication via language takes place. It is firmly rooted in the "believe that [...] we must consider language ability within an interactional framework of language in use" (Bachman & Palmer 1996: 78). The Bachman framework combines language knowledge and a number of factors related to the real-world context of communication in a very detailed and comprehensive way. As the model clearly implies, the different components of CLA cannot be seen in isolation from each other but rather as complementary components of language proficiency as a whole. This is not only the case for the three broad components of CLA but also for their respective sub-components. If one takes a closer look at language competence, for instance, it becomes clear that all the sub-components of this category need to operate in close connection with each other in order for language to be processed successfully. The first component of language competence, organizational competence, "which includes *grammatical knowledge* and *textual knowledge*, enables language users to create and interpret utterances or sentences that are grammatically accurate, and to combine these to form texts, either oral or written that are cohesive and rhetorically or conversationally organized" (Bachman & Palmer 1996: 78). It would, however, not have a very significant value if seen in isolation from pragmatic competence, the second major component of language

competence. Pragmatic competence "enables language users to relate words, utterances, and texts to concepts, communicative goals, and the features of the language use setting" (ibid.).

Looking at the interrelations just mentioned from a grammar-based point of view, it becomes obvious that grammatical competence, i.e. "the knowledge of vocabulary, morphology, syntax, and phonology/graphology" (Bachman 1990: 87) needs to be combined with textual competencies in order to be able to form and interpret grammatically correct utterances that exceed the sentence level. Moreover, grammatical and textual competence, even if seen in close relation to one another are not very useful with regard to actual communication unless they become related to a number of factors such as pragmatic (i.e. illocutionary and sociolinguistic) competencies, as well as strategic and psychophysiological processes involved in language use.

This view of a high interrelatedness of grammar and other aspects of language knowledge and use is also reflected in present day approaches to grammar testing. Such approaches have already been outlined in chapter 3.1, they shall, however, be revised briefly at this point, as most of them can be seen as clear results to changes in the definition of language proficiency just discussed:

Firstly, the explicit testing of grammatical knowledge in isolation from other language competencies has lost influence in many EFL contexts (c.f. Rea-Dickins 2001: 23; 29). Secondly, there is nowadays a relatively broad definition of the grammar concept notable (c.f. Rea-Dickins 2001: 28). This definition can be described as a combination of Bachman's grammatical and textual competence categories. Thirdly, even techniques or formats used in explicit grammar testing show a tendency towards an integration of grammar and a number of other language skills (ibid.), which is one major characteristic of the ability to communicate in real-life contexts.

As shall be outlined in closer detail in the course of this paper, both the WBPT and the grammar sections of the TUC-PT put a main emphasis on grammatical and to a certain extent textual competence. The analysis of the two formats will thus primarily focus on the grammatical and to a certain extent also the textual structures that are being tested. At the same time, however, a closer look will be taken at the potential of the different task formats and item types for integrating certain aspects that lie outside the domain of grammar competence. Particularly, aspects of the amount of (authentic) language context that is provided for each item or questions of testing grammar within test formats that are typically used for the assessment of overall language proficiency will be part of the discussion. Before analysing the two test types with reference to task

formats and item types, both of them will be formally described and compared on the basis of the test classification framework provided by Bachman (1990). The following chapters centre around a detailed description of the Bachman framework and its application to the two tests to be analysed and compared.

4. Formal Description of the WBPT and the TUC-PT Based on Bachman (1990)

As mentioned earlier on, the Bachman framework for test classification consists of "five distinctive features" (1990: 70). In the following, these features will be explained in a more detailed way and will then be applied to the TUC-PT and the WBPT. As the WBPT was designed with the primary purpose of serving as a practise opportunity for the TUC-PT a number of formal characteristics of the former can be described as direct results of its close orientation to the latter. Thus, the Bachman framework will first be applied to the TUC-PT and in a second step, the findings will be compared with what can be found out to be the "distinctive features" (ibid.) of the WBPT.

4.1 Intended Test Use

Bachman looks at the intended use of a language test primarily from an educational point of view, namely with reference to the "decisions to be made" on the basis of a given test (1990: 70). He thus arrives at three groups of language test types, each of them being based on a particular type of decision. The first group of tests are used for making "admissions decisions" (ibid.), i.e. decisions about whether or not a test person will be admitted to an educational programme, a profession, a university, or even a country etc. Types of tests which are specifically used with such an intention are, according to Bachman, "*selection*, *entrance*, and *readiness* tests" (ibid.).[15] What is noteworthy about these types of tests, at least with reference to selection and/or entrance tests, is the fact that their outcomes "generally involve major consequences for test

---

[15]The boundaries between what characterizes these three types of tests do not seem to be very clear-cut. The terms *selection* and *entrance* test, for instance, are sometimes used synonymously in the literature (c.f. Davies et al. 1999: 66), and I shall not go into a detailed discussion of the similarities and differences of the three tests types just named. For a more extended discussion of the test types, c.f. Bachman 1990: 58; 70f.

takers" (Davies et al. 1999: 66) which may affect their personal, educational, and/or professional future in a more or less drastic way.[16]

A second use of language tests can lead to decisions about the "instructional level" that is appropriate for a test taker or about "specific areas in which instruction is needed" (Bachman 1990: 70). Test types which are typically used as a basis for such decisions are *placement* and *diagnostic* tests (ibid.). In contrast to the group of test types mentioned above, placement as well as diagnostic tests are not used for deciding whether or not a test taker will be admitted to a certain teaching programme etc. In fact, the primary purpose of diagnostic tests is to decide which areas of language study are especially problematic for a given test taker. They often function as a basis for providing language learners with "further help" with their learning (Alderson et al. 1995: 12). Like diagnostic tests, placement tests, as well as any other type of language test, can provide information about those areas of a language that might be problematic for language learners. The primary use of a placement test, however, is to determine which level of language instruction is the most suitable for a given test taker; and this level is typically determined before the instruction starts. Obviously, "it is vital that the information gained from test scores is appropriate for the decisions that will be based on them" (Fulcher, 1999: 290). In order to be valid, a placement test should therefore "reflect the features of the teaching context" (Davies et al. 1999: 145) into which the test taker is to be placed. Such features can be related to "the proficiency level of the classes, the methodology and the syllabus type" (ibid.). A clear orientation towards a certain type of teaching context is one characteristic that placement tests share with the third group of tests that are classified by Bachman according to their intended use. This group consists of "*progress*, *achievement*, *attainment*, or *mastery* tests". With respect to their intended use all these test types have in common that their outcomes can be used for deciding "how individuals should proceed through the program, or how well they are attaining the program's objectives". (Bachman 1990: 70). Although placement as well as this third group of tests entail or should entail a certain focus on a given teaching context, for the group of *progress*, *achievement*, *attainment*, or *mastery* tests, this focus is more powerfully restricted to the actual course content or syllabus.[17] Whereas for *placement* tests there exists a variety of ways for reflecting the teaching context to

---

[16]This does not seem to be of primary importance as far as readiness tests are concerned. Their purpose is first of all "to determine whether or not" a test person is "ready for instruction" (Bachman 1990: 58)

[17] *Progress* tests, for instance, are intended to measure an individual's language ability "at various stages of a language course" in order to assess that person's language learning progress (Alderson et al. 1995: 12). In contrast to that *achievement* tests are not so much concerned with the learning progress as such but with the actual outcomes of the learning process. They can be described as instruments "designed to measure what a person has learned within or up to a given time" and their "content is based on a sample of what has been in the syllabus during the time under scrutiny" (Davies et al. 1999: 2).

follow, *progress*, *achievement*, *attainment*, or *mastery* tests need to show a relatively explicit orientation towards a certain kind of course syllabus. Furthermore, there exist certain differences between placement tests and the group of tests just mentioned as far as the point of time of their administration is concerned.

Applying the different features of the use or purpose a test can be intended for and the types of decisions that can be taken on the basis of a test to the two tests that are to be analysed and compared by this paper, it becomes obvious that the TUC-PT can clearly be described as a placement test. Its primary use is to determine at which stage in the language instruction programme of the English practical language programme at Chemnitz University of Technology a candidate is to be placed or, in other words, which type of language course 1st semester students of English and American Studies or one of the related subjects will start their practical language instruction with. The TUC-PT serves as a means for deciding on a placement of students into three language courses: these courses are called Foundation Course, Integrated Language Course I (ILC I), and Integrated Language Course II (ILC II). The three have ascending difficulty levels and relatively diverse focuses as far as the course content is concerned. The Foundation Course puts a main emphasis on a revision and extension of students' "abilities in essential grammar and vocabulary areas" with a focus on oral as well as written English language skills (*Chemnitz University of Technology English and American Studies Kommentiertes Vorlesungsverzeichnis Wintersemester 2004/05*: 26). In contrast to that, the primary focus of ILC II does not explicitly lie upon grammar and vocabulary but on productive language skills. In this context, a special importance is attached to types of such skills "required for a broad spectrum of professional fields with the relevant text" and "speech functions, and lexical areas linked to specific objective-related topics". Skills focused on by ILC I are for example: "letter-writing (formal/informal), summary & report, review & assessment, role play, informal discussion" and "more formal debate" (ibid.). A relatively high proportion of the course content aims at a preparation of students so that they become able to use the English language effectively in university settings, as can be seen for example with regard to the importance of report or summary writing or the ability to actively take part in more or less formal discussions and debates. At this stage of the practical language instruction programme the significance of explicit grammar teaching does not seem to be as high as in the Foundation Course. Rather, explicit grammar practise is more or less restricted to self-study (ibid.). Comparing the course contents of the Foundation Course and ILC I it becomes obvious that the latter puts a main emphasis on skills used for successful communication – inside and outside university settings – with grammar and vocabulary

knowledge being treated as implicit components of communicational skills rather than as explicit contents of a language course as is the case for the Foundation Course. This goes along with a general tendency that can be observed, according to Rea-Dickins and a number of other researchers, in many EAP contexts. She points out that in EAP, "grammar appears to assume a less prominent and explicit role at later stages in the instructional programme" (Rea-Dickins 2001: 29). This tendency is continued in ILC II which aims at an extension of the principles of ILC I with a special emphasis on the use of English in communicative situations and skills such as "expressing views, structuring information-communication and arguments, interrupting, clarifying etc." (KoVo WS 2004/05: 27). Questions of appropriate communicative reactions as well as the use of adequate register for different communicational situations play a crucial role in ILC II (ibid.). Normally, most of the test takers of the TUC-PT are either placed in the Foundation Course or in ILC I, there exist however a few cases where test takers with exceptionally high language abilities are allowed to start the practical language instruction with ILC II.

The TUC-PT can thus be described as an instrument for deciding on a placement of first semester students either in a Foundation Course, into ILC I or (in some rare cases) into ILC II. Furthermore, the TUC-PT shares one of the very obvious characteristics of placement tests as it is administered before the language instruction programme starts.

Comparing the TUC-PT with the WBPT it becomes clear that the latter does not share all of the features that are typical for placement tests. Strictly speaking, the primary characteristic that distinguishes placement tests from other test types is not a feature of the WBPT; i.e. there is no decision about where to place test takers in a language course based upon the outcomes of the WBPT. Moreover, it is impossible to identify the WBPT as a member of the other two groups of tests, Bachman distinguishes with reference to the intended use of a test. In fact, the primary intention the WBPT was designed with is to serve as an opportunity to practise for the grammar sections of the TUC-PT. It is thus relatively problematic to describe the position of the WBPT within Bachman's proposed framework of intended test uses. Nevertheless, especially as far as the grammatical contents of the relevant test sections is concerned, the test was designed with a strong orientation towards a test format which is clearly identifiable as a placement test. It is for that reason as well as for reasons of conciseness that I decided to refer to the WBPT as a placement test. It needs, however, to be clearly emphasised here that the WBPT does not only lack a the typical characteristic of a placement test with reference to its intended use, there are also a number of other characteristics that

distinguish it from most language placement tests in general and the TUC-PT in particular. Such characteristics concern for example the testing situation and include factors such as the supervision by a language teacher, tester or any other kind of personnel. This means that although test takers have to finish the WBPT within a certain time limit, there is the potential possibility to look up unfamiliar language in (on-line) dictionaries, grammar reference works and the like. Furthermore, it needs to be underscored at this point that the WBPT only focuses on grammar and grammar terminology, i.e. on a restricted selection of the skills measured by the TUC-PT.

4.2 Test Content

As has been mentioned earlier on, when referring to the content of a language test, Bachman distinguishes "theory-based" from "syllabus-based tests (1990: 71). The two prototypical test types for these categories are *proficiency tests* and or *achievement tests,* respectively.[18] The following discussion of syllabus- and theory-based tests will thus be provided on the basis of the differentiations that can be made between achievement and proficiency tests as the prototypical members of the two categories. It is very important to keep in mind that the distinction that exists between theory- and syllabus-based tests, however, does not imply that the former do in any case measure language abilities that are completely different from those measured by the latter, or as Bachman puts it:

> Whether or not the abilities measured by a given proficiency test actually differ from those measured by a given achievement test will depend, of course, on the extent to which the theory upon which the proficiency test is based differs from that upon which the syllabus is based. For example, a proficiency test based on a theory of grammatical competence is likely to be quite similar to an achievement test based on a grammar-based syllabus (ibid.).

Despite such possible overlaps there exist a number of differences between proficiency and achievement tests. Davies, for instance, draws a very clear distinction between achievement and proficiency tests. For him one of the main differences between the two test types seems to consist in the question whether or not the test is preceded by a certain kind of shared learning process. He points out that

> Achievement tests are [...] based on a clear and public indication of the instruction that has been given. Proficiency, on the other hand is concerned not with publicly stated instruction but with the relationship between language control and a particular use of language, for example, with whether a testee has adequate language for academic study (1990: 6f.)

---

[18]Bachman further mentions *language aptitude tests* as another example for theory-based tests. What distinguishes this group of tests from proficiency tests is the fact that "the theory upon which they are based includes abilities that are related to the acquisition, rather than the use of language" (Bachman 1990: 72).

Like Davies, Lado refers to the shared "learning experience" that achievement tests are based on as the most important characteristic that distinguishes them from proficiency tests which "measure how much of a foreign language a person [...] knows" (1961: 369) without much emphasis being put on the type of instruction such knowledge originates from. Almost 35 years later this assumption is still reflected in the different ways of defining proficiency tests and their relation to other test types, especially achievement tests. However, whereas Lado gives a quite general definition of what proficiency tests are intended to measure, present day research tends to give more specific examples of the potential uses of proficiency test scores in real life situations. This fact is not only evident in Davies' description of the category but also in a definition of the term *proficiency test* provided by Alderson et al.:

> *Proficiency tests* [...] are not based on a particular language programme. They are designed to test the ability of students with different language training backgrounds. Some proficiency tests [...] are intended to show whether students have reached a given level of general language ability. Others are designed to show whether students have sufficient ability to be able to use a language in some specific area such as medicine, tourism or academic study" (1995: 12).[19]

Measuring a test person's level of language ability, whether general or specific, is not only a defining characteristic of proficiency but also of most placement tests which aim at measuring such levels with the primary intention to place testees at a particular stage in a language instruction programme that appropriately reflects his/her level of language ability. With reference to the principles that placement tests are based on, or in Bachman's words, to their content, Wall et al. observe that

> typically, placement tests are of two types: the first has a proficiency orientation. [...] The content of such tests bears no direct relationship to the content of the language courses on to which students are being placed. The second type is intended to reflect the nature of such language courses, and might be said to be 'preachievement' (1994: 322).

Referring to the possible "proficiency orientation" of placement tests, they clearly establish a relation to Bachman's category of theory-based tests.[20] Focusing on the second type of placement tests that is mentioned by Wall et al. it becomes possible to

---

[19]The fact that descriptions and definitions of proficiency tests have become more and more detailed with a special emphasis being put on the possible uses of proficiency test scores, i.e. the inferences that can be drawn from the results about the ability of a test person to cope with general and/or professional language demands points to a growing interest in proficiency testing since the 1960s. This is also reflected in the high importance of standardized proficiency exams of English such as the TOEFL or the Cambridge exams. This can be regarded as an indicator of the importance of English as an International Language in present-day globalised societies.

[20]Moreover, they even go so far as to suggest that a placement test "may indeed be an institutionalised version of a proficiency test like the [...] TOEFL" (ibid.) or, in other words that proficiency tests can be used for placement purposes. Similarly, Davies et al. describe placement as one possible purpose a proficiency test can have (1999: 205).

suggest that this type reflects Bachman's category of syllabus-based tests because there is a clear orientation towards a certain type of syllabus notable for this type.[21] The remainder of this chapter, will take a closer look at the principles that underlie the WBPT and the TUC-PT with special reference to the question whether the two tests can be described as theory- or rather syllabus-based types of tests.

Both, the TUC-PT and the WBPT are not explicitly based on a certain kind of course syllabus. As the WBPT was designed in very close relation to sections of the TUC-PT the content orientation of both tests is relatively similar. The TUC-PT in its entirety aims at measuring test takers' state of language proficiency at the moment when the test is delivered (Petra Naumann, personal communication, 11 Nov. 2004). As has been outlined in chapter 3.3 above, the concept of language proficiency has been described in a variety of ways. What most present-day definitions of language proficiency have in common, however, is a strong emphasis on actual language use. This can be seen, for instance, in the definition of the concept provided by Davies et al. according to which language proficiency is "a general type of **knowledge** of or **competence** in the use of a language" (1999: 153). Accordingly, in his definition of CLA, Bachman focuses very clearly focuses on the ability to implement knowledge of a language "in appropriate, contextualized communicative language use" (1990: 84). Looking at the TUC-PT as a whole, it becomes notable that the test does not only focus on the components of test takers' language competence, such as grammar and vocabulary competence or the knowledge of spelling rules but also on their ability to transfer such knowledge to different oral and written tasks, such as giving personal information and interacting in an oral interview or letter writing. These tasks clearly do not only focus on rather passive language knowledge: The oral interview, for example focuses on an integration of the skills of listening and speaking, on changing roles between speaker and hearer and so on which can be considered to be defining characteristics of natural real-life communication in its oral form. Furthermore, the interview normally requires test takers to give personal information about themselves, their family, their plans for the future etc., i.e. the interview focuses on situations and experiences that are part of his/her actual life and thus contains a relatively high amount of authenticity. Similar to the oral interview, the writing task requires an integration of several skills such as grammatical, lexical or orthographical competence which certainly is an important feature of written communication in real-life contexts. Nevertheless, it needs to be pointed out that writing a letter to an imaginary pen friend will render the

---

[21]The fact that Wall et al. speak of a "'preachievement'" (1994: 322) rather than an achievement orientation is simply a result of the general characteristic of placement tests being administered before a language instruction programme starts.

task relatively artificial for most test takers. The writing section of the TUC-PT, therefore, seems to contain a relatively small extent of authenticity when compared to the oral interview. However, it needs to be maintained that writing a letter involves a focus on the ability to write for an addressee which is an important feature of many forms of written communication. To sum up, the writing section as well as the oral part of the TUC-PT put a more or less high emphasis on test takers' ability to use the English language in communicative contexts, and can therefore be said to reflect at least some of the principles that underlie present-day approaches to language proficiency, such as Bachman's CLA model.

In contrast to those parts of the TUC-PT just discussed, for the spelling and the grammar sections of the TUC-PT the main emphasis is on testing knowledge as such without much reference to the use of this knowledge in actual communication. As far as the grammar sections are concerned the tasks are more discrete-point than those that are part of the other sub-sections. Most of the items focus on one individual grammar point, respectively and they seem to be in accordance with the following definition of discrete-point testing provided by Oller:

> Traditionally, a *discrete point* test is one that attempts to focus attention on one point of grammar at a time. Each test item is aimed at one and only one element of a particular component of a grammar [...], such as phonology, syntax, or vocabulary. Moreover, a discrete point test purports to assess only one skill at a time (e.g. listening, or speaking, or reading, or writing) and only one aspect of a skill (e.g. productive versus receptive or oral versus visual). (1979: 37).

Applying Bachman's CLA model to the TUC-PT it becomes possible to suggest that its grammar sections rather exclusively focus on those aspects of language proficiency that Bachman labels organizational competence and which "comprises those abilities involved in controlling the formal structure of language for producing or recognizing grammatically correct sentences, comprehending their propositional content, and ordering them to form texts" (1990: 87). The most important component of this organizational competence that is measured by the TUC-PT is grammatical competence although there is some focus on textual competence notable as well. In contrast to that, the writing and especially the oral section of the test also involve a clear focus on factors such as pragmatic as well as strategic competence in addition to measuring grammatical (and textual) competencies.[22]

Although the grammar sections of the TUC-PT aim at the measurement of a limited number of components of language proficiency or CLA, it can be assumed that

---

[22] Like all the other sub-sections of the TUC-PT, the grammar sections require certain psychophysiological mechanisms in order for test takers to complete the respective tasks successfully. As these mechanisms are not part of the present analysis, they will not be discussed in further detail in this paper.

the test as a whole measures a variety of aspects related to the concept, i.e. it includes aspects of language proficiency that go well beyond the level of mere language knowledge. The grammar sections as well as the whole test cannot be described as syllabus-based as they are not directly based on "a specific domain of content" (Bachman 1990: 71). Of course, there needs to exist a certain relation between a placement test and the language course(s) following that test, and this is also true for the TUC-PT. However, there exists no clear documentation of the relation between the test and the contents of the instructional programme to follow. Rather, the main emphasis of the test lies upon a variety of components of language proficiency in general. Applying the two types of placement tests distinguished by Wall et al. to the TUC-PT it, thus, becomes obvious that the test has a proficiency rather than a preachievement orientation. With reference to the distinction between theory- and syllabus-based tests, it is, therefore, possible to suggest that the content orientation of the TUC-PT is theory- rather than syllabus-based. [23]

As far, as the grammar sections of the TUC-PT are concerned, it is evident that they focus on grammar topics which can be assumed to be part of the grammatical competence of a student at 'Abitur'-level and it can be said that although the target group of the TUC-PT does not entirely consist of German students the German 'Abitur' does to a certain extent function as a basis for deciding on the topics of the grammar sections of the test. Accordingly, these topics include: tense and aspect, the use of adjectives and adverbs, gerund and infinitive constructions, conditional clauses, the use of prepositions etc. as well as the terminology of English grammar (Petra Naumann, email to K.U., July 8, 2004). Nevertheless, it needs to be maintained that these contents are not directly derived from an explicitly stated syllabus, rather they need to be viewed as one aspect of the level of language proficiency students at 'Abitur'-level are generally assumed to have. Although most of the test takers have finished their school education

---

[23]It is nevertheless possible to ascribe certain aspects of a syllabus orientation to the TUC-PT. Like an achievement test, the TUC-PT focuses on a certain spectrum of language competencies and abilities that have been acquired on the basis of a more or less clearly defined instructional programme, in this case the level of language proficiency typical for test takers whose educational level is the German 'Abitur'. At this level, a test person can be assumed to possess a number of relatively specific competencies, such as the reflection of various types of texts and of language as such, the use of different presentation and interaction strategies etc. (*Lehrplan Gymnasium Englisch Klassenstufen 5 bis 10, Jahrgangsstufen 11 bis 12*: 68-82*)*, the basis for such abilities seems, however, to be the ability to use the English language in a variety of contexts which is strongly in accordance with the definition of language proficiency provided by Davies et al. (c.f. 1999: 153) and also with Bachman's definition of CLA (c.f. e.g.1990: 84). The Saxon Ministry of Culture points out: "Im Englischunterricht erwerben die Schüler situativ angemessene differenzierte Kommunikations- und Diskursfähigkeit in mündlicher und schriftlicher Form" (*Lehrplan Gymnasium Englisch Klassenstufen 5 bis 10, Jahrgangsstufen 11 bis 12*: 15) and, thus, points to the significance of the ability to use English in a very broad spectrum of contexts which students should possess when finishing their 'Abitur'. The TUC-PT seems to take this broad English language proficiency that students should possess at 'Abitur' level as a basis, rather than basing measurement on the concrete content domains that are part of the respective syllabuses.

with the German 'Abitur', some of them come from countries other than Germany and might therefore have different educational backgrounds. Others might already have finished a certain course of studies or parts of it. Such factors render the target group more or less heterogeneous and, thus, make it impossible to directly base the content of a test on one explicitly stated syllabus. Summarizing, the TUC-PT can be described as an example of theory-based tests in the sense of Bachman (1990), it is not explicitly based on a course syllabus, neither that of the preceding language instruction programme nor that of the language course(s) to follow.

Whereas the TUC-PT in its entirety can be described as focusing on a combination of different components of language proficiency, the WBPT does not claim to be a reflection of a comprehensive theory of the concept, such as the one provided by Bachman (1990). It needs to be taken into account that the WBPT only focuses on reflecting some of the sections of the TUC-PT. These are first of all grammar or grammatical and textual competence. The WBPT can thus be described as theory-based in a restricted sense, i.e. a measure of the components of language proficiency mentioned above. Like the TUC-PT, the WBPT focuses on a variety of grammatical topics such as tenses, prepositions, relative clause constructions, the use of adjectives versus adverbs, conjunctions, quantifiers, etc. Like the TUC-PT, it is, however, not based on an explicitly stated syllabus neither of the language course preceding the instructional programme nor the one to follow it and does thus not have an achievement nor a preachievement orientation.

4.3 Frame of Reference

According to Bachman, "the results of language tests can be interpreted in two different ways, depending on the frame of reference adopted" (1990: 72). Language tests can either be norm-referenced (NR) or criterion-referenced (CR) (ibid.). For NR tests "scores are [normally, K.U.] interpreted in relation to the performance of a particular group of individuals" or a so-called norm-group (ibid.).[24] They can thus be described as providing "information about the relative ability of candidates" (Wharton, n.d) when compared to the norm-group. This seems to be a crucial characteristic that distinguishes NR from CR tests. The latter do not measure a test taker's language ability in relation to that of other test takers but in relation to a particular "level of **knowledge** of, or **performance** on, a specific domain of target behaviours (i.e. the **criterion**) which the **candidate** is required to have mastered" (Davies et al. 1999: 38). A typical example of a CR test is, according to Bachman, "the case in which students are evaluated in

---

[24]Standardized tests, e.g. the TOEFL; are often referred to as prototypical examples for NR-tests, for a more detailed discussion of norm-referencing as well as standardized tests, c.f. Bachman 1990: 72f.

terms of their relative degree of mastery of course content, rather than with respect to their relative ranking in class" (1990: 74).

The results of both the TUC-PT and the WBPT are not interpreted in relation to the results of another group of test takers on the same test. Rather, test scores are used as a basis for deciding about the candidates' mastery of different components of the English language as well as different competencies that can be described as being part of English language proficiency as such. Both tests can therefore be described as CR rather than NR tests, although it needs to be kept in mind "that these two frames of reference are not necessarily mutually exclusive" (Bachman 1990: 76).

Most importantly, a CR test needs to be developed in such a way "that it adequately represent the criterion ability level or sample of the content domain and that it be sensitive to levels of ability or degrees of mastery of the different components of that domain" (Bachman 1990: 74). Accordingly, the levels of language ability or of mastery of certain features or components of a language seem to be most important for this feature of language tests. As the present paper is most interested in placement tests their content and the methods used in such tests, the aspect of frame of reference will not be further discussed here. It needs to be kept in mind, however, that for a complete description of language tests on the basis of the framework provided by Bachman (1990), all of his classification features need to be taken into account, or as Bachman himself puts it "descriptions of language tests that refer to only a single feature are likely to be misunderstood" (1990: 77). The respective amount of emphasis that is put upon the individual description categories depends, however, on the main focus of the analysis and the relevance of the individual features for such an analysis.

4.4 Scoring Procedure

Referring to the possible procedures that can be used for scoring a language test, Bachman distinguishes the two main categories of objective and subjective scoring. The terms *objective* and *subjective test* as well as *objective* and *subjective* scoring are often used interchangeably in the literature and will also be used in such a way in the present paper. It needs, however, to be maintained that "it is not really the test that is objective [or subjective; K.U.] but the systems by which they are marked" (Power 2003) The main characteristic that distinguishes subjective from objective tests is the fact that "the scorer must make a judgement about the correctness of the response based on her subjective interpretation of the scoring criteria" (Bachman 1990: 76). Typical examples for tests which require subjective scoring are "the oral interview or the written

composition" which often "involve the use of rating scales"; for such tests "there is no feasible way to 'objectify' the scoring procedure" (ibid.)

In contrast to subjectively scored forms of assessment, for objective tests the "correct responses are clearly specified, and markers are not required to make judgements" (Davies et al. 1999: 133). This implies that in an objective test an answer can only be correct or incorrect, whereas subjective tests often require the scorer to make decisions about degrees of correctness. The most common testing techniques or formats used in objectively scored tests are, for example, multiple choice or exact cloze formats as well as dictations (c.f. Power 2003; Bachman 1990: 76). It is important that for all of these techniques in order to be suitable for objective scoring, a clear identification of the correct answer(s) or solution(s) must be given, this is generally done in the form of an answer key (Bachman 1990: 76).

Since comprehensive and fixed answer keys for objective test tasks are relatively easy to establish, such tasks seem to be predestined to be scored by machine. In fact, the multiple choice technique as one of the most common forms of objective testing and a number of other objective test formats such as exact cloze are very frequently used in computer-based language tests and particularly often in computer-based tests of grammar. Roever points out that objective testing techniques such as the ones just mentioned are "fairly easy to implement" in computer- or web-based language testing contexts (2001: 86). Moreover - and this is probably one of the most important reasons for the ever growing importance of computers in language testing – technology can help to economize and facilitate the scoring process, especially if there exists only a limited number and a comprehensive key of possible correct responses. Besides these obvious advantages of objective test techniques, they also have quite a number of serious drawbacks, particularly when a language test consists entirely of objective item types. Many types of objective testing, especially the multiple choice technique have often "been criticised for trivialising assessment and for being amenable for guessing" (Davies et al. 1999: 132). As objective test items often require the test taker to chose from a predetermined list of responses or to complete missing parts of an already existing language context the possibilities for using a language freely and creatively are rather restricted in such tasks. Accordingly, objectively measurable testing techniques have often been criticised for not requiring the test taker to actively produce language in his/her response. Rather, objective testing largely focuses on receptive processing of language input (Brown 1997: 45). Likewise, the high frequency of objectively measurable testing formats, especially multiple choice items, in present-day computer-based language testing contexts has been a target of criticism for many researchers. In

this respect, Alderson and Banerjee refer to Chalhoub-Deville and Deville (1999) whose view of the problem they summarize in the following way.

> Chalhoub-Deville and Deville (1999) point out that, despite the apparent advantages of computer-based tests, computer-based testing relies overwhelmingly on selected response (typically multiple choice questions) discrete-point tasks rather than performance-based items, and thus computer-based testing may be restricted to testing linguistic knowledge rather than communicative skills (Alderson & Banerjee 2001: 225).

Being more or less restricted to the measurement of receptive skills and language competence rather than performance can be described as one of the crucial drawbacks of objective testing formats, no matter if they are used in computer-based or paper-and-pencil testing. Spolsky even goes so far as to suggest that "purely objective testing appears to have failed" and furthermore points out that, although the results of purely objective tests can be assumed to be very reliable such tests do not actually measure language proficiency in the sense of an ability to use language in a variety of contexts (1995: 349). Apart from all the criticisms that have been made about objective language testing it needs to be stated that such tests also provide a number of advantages. Davies et al., for instance, point out that objective tests have been proven by empirical research to be able to "provide broad subject-matter coverage" and that their scores are "more reliable and [...] fairer" than those of subjectively scored tests (1999: 132). Summarising, it can be assumed that both objective and subjective scoring procedures have their advantages as well as their drawbacks. In general, it can be assumed that in order for a language test to provide a reliable and valid account of test takers' overall language ability, a combination of subjective and objective techniques might prove useful. This is probably one of the reasons why many standardized language tests, such as the TOEFL or many of the language examinations by the University of Cambridge Local Examinations Syndicate (UCLES), combine subjective testing, such as written composition tasks and oral interviews with objective ways of testing language structure (in most cases grammar and vocabulary).

Like most of these examinations, the TUC-PT also combines subjective and objective testing techniques. The grammar sections are entirely multiple choice which means that for each test item there is a limited number (between three and four) of possible responses given from which the test taker has to select the correct one. For all these sections there exists a fixed answer key which is strictly adhered to by the test scorers. For the spelling section of the TUC-PT there exists an objective key as well. As this section, however, requires test takers to correct mistakes in a number of sentences without any specifications concerning the number or type of mistakes being made, there is always the possibility for test takers to come up with a solution which is not part of

the key but which is nevertheless acceptable. Thus, the spelling section can be said to involve a certain amount of subjective scoring whereas the grammar sections are scored purely objectively. Obviously, the writing and the oral component of the TUC-PT require subjective scoring methods as both sections focus on a more or less free and creative use of language which cannot be assessed with the help of a fixed answer key.[25]

Summarising, the TUC-PT combines subjective and objective scoring methods, however, subjective judgements by experienced language teachers and testers, seem to have a more important significance for the final placement decision to be taken than scores in the objective part of the test. Nevertheless, test scores in the spelling and grammar sections, which are largely gained by objective scoring, can be said to form a preliminary basis for the actual placement of candidates, which can then be adjusted according to subjective judgements made about him/her on the basis of his/her performance in the writing and speaking sections of the test. A sore of 60% in the spelling and grammar sections is considered to be the 'pass mark'. Candidates with a score of less than 60% in these sections are automatically placed into a Foundation Course and only an exceptionally good performance in the written and/or spoken section might lead to a change of this decision. Accordingly, the objectively measurable parts of the test seem to play a crucial role for the decisions that are made on the basis of it. This does, however, not seem to imply that objective testing methods are preferred over subjective ones, but that there is a relatively high importance attached to grammatical competence, and thus to test taker scores in the grammar sections, by the TUC-PT.

The fact of the WBPT being administered in a computer- or web-based format, can be viewed as one of the reasons for the whole test being objectively measurable. As it is intended to be a practise opportunity for the grammar sections of the TUC-PT which is to be scored quickly and effectively, the WBPT follows the general tendency of computer-based tests of grammar frequently being objective in nature. For each item in the WBPT there exists a limited number of correct alternatives, in general one correct answer to each item. Accordingly, the final test score each test taker is provided with once the test is finished, is gained by a strict adherence to the answer keys for the respective tasks and subjective judgements about test taker responses cannot be taken into account. This, of course, entails a number of implications for task and item design, such as the selection of genuinely incorrect alternatives in multiple choice formats etc.

---

[25]Furthermore, especially the oral interview serves as a means for deciding on the appropriate instructional level for a given test taker not only on the basis of his/her language ability but also with reference to personal characteristics such as the ability to adopt to the special demands of learning English for Academic Purposes (Petra Naumann, personal communication, November 11, 2004).

For the different multiple choice tasks there exists a definite answer key in which the respective number of the correct alternative is recorded. The C-Test component, as the only section of the WBPT which does not employ a multiple choice format, requires test takers to supply the missing half of particular words in a text by typing them in. Accordingly, the answer key for the C-Test contains a list of the correct letter combinations for each gap. One of the reasons for designing the WBPT in an objective format was the fact that it can easily be scored on the computer which can help economize the scoring process in a crucial way. This was done with a special view to whether it would be possible to deliver certain sections of the TUC-PT via the computer in the future. Accordingly, the WBPT is not only a practise version for the grammar sections of the paper and pencil placement test, it can also be viewed as an experimental test version used to find out how feasible it would be to convert certain sections of the TUC-PT, at least those that are objectively measurable, into a computerized format.

Keeping in mind that entirely objective tests can only be used to assess restricted areas of language proficiency or ability, it needs to be clearly outlined again that the WBPT has to be viewed as a practise opportunity for just one part of a broader language test battery and, thus, does not claim to be a measure of all aspects of language proficiency. The WBPT mainly focuses on grammatical as well as certain aspects of textual competence which are according to Bachman (1990) an important part of language proficiency or the ability to actually use a language. However, in order for a language test to be a valid measure of this ability, other testing methods as well as scoring procedures need to be taken into account. This is partly done by the subjective parts of the TUC-PT.

4.5 Test Method

The fifth and final characteristic Bachman proposes for the classification and description of language tests is the test method. He points out that "given the variety of methods that have been and continue to be devised, and the creativity of test developers, it is not possible to make an exhaustive list of the methods used for language tests" (1990: 77). Nevertheless he names some "commonly used methods, such as the multiple-choice, completion (fill-in), dictation, and cloze" technique (ibid.). He further points out that these techniques are not sufficiently described by applying a label like the ones cited above to them; rather, all of them "consist of different combinations of features" or "test method 'facets'" (ibid.). In *Fundamental Considerations in Language Testing*, Bachman provides a very detailed and comprehensive list of possible test method facets. It needs to be taken into account that all of these facets seem to have a

very high significance for a given language testing process as they are assumed to have a very strong effect on a candidate's performance on a test or test task (c.f. Bachman 1990: 112f.).[26] Comparing a language testing situation to real-life language use contexts, it becomes possible to view the facets of the test methods as "restricted or controlled versions" of "contextual features" that determine real-life language performance or use (c.f. Bachman 1990: 111f.). As the following part of the present paper will put a special emphasis on the analysis of linguistic aspects of different grammar testing techniques and items and not so much on the contextual features influencing language performance on particular test tasks, the test method facets established by Bachman will only be mentioned and discussed briefly here.[27] Nevertheless, the impact of different test methods on language test performance should be kept clearly in mind and the different methods used by the TUC-PT and the WBPT will be discussed in closer detail in the course of this paper.

With reference to the methods that can be used in language tests, Bachman distinguishes the following categories of features or facets of such methods: The first broad group of such features is concerned with a number of "facets of the testing environment" (1990: 119) which include factors such as "familiarity of the place and equipment used in administering the test; [...] the personnel involved in the test" etc. (118). The second group comprises "facets of the test rubric"(119) such as "test organization, time allocation, and instructions" (118). A third of Bachman's categories is concerned with the "input and expected response" provided and required in different test methods. Facets of this category include the format of input and expected response, and a number of sub-facets, such as "channel and mode", "form", "vehicle" and "language of presentation" of the input and the "type", "form" and "language of expected response" (127-130). Furthermore, the third category of test method facets includes a number of features related to the "nature of language input and expected response", these are for instance length and propositional content (c.f. 130-139). Moreover, the author mentions "organizational characteristics" which "are related to [...] the formal organization of [...] the discourse of both the input and the response" (139). He distinguishes "three types" of organizational characteristics, these are "grammar, cohesion and rhetorical organization" (ibid.)[28]. The "nature of input and

---

[26]This relation between the test method and test performance is also reflected in the notion of the so-called *test method effect* which is generally judged as a negative side effect of language tests as "the method used for testing a language ability may itself affect the student's score" (Alderson et al. 1995: 44) in such a way that the score does not provide valid information about a test person's language ability as such but rather about his/her ability to handle certain testing techniques or methods.
[27]For a more detailed account of Bachman's classification of test method facets, c.f. 1990: 116-152
[28]This division can be described as being directly related to organizational competence as one component of CLA (c.f. Bachman 1990: 87-89; 139).

expected response" (130) is also affected by "pragmatic characteristics" which include facets such as "illocutionary force" and a number of "sociolinguistic characteristics" (c.f. 140-144). Within a fourth broad category of test method facets, Bachman mentions different "restrictions on expected response" (144). An example for this would be the "highly restricted" response format of selected response methods (145). A final group of test method facets is concerned with the "relationship between input and response", these can either be reciprocal, non-reciprocal or adaptive (148).[29]. The different test method facets established by Bachman are outlined in Appendix I.

As the following sections of this paper will be concerned with an analysis of the different test methods and item types used in the WBPT and the grammar sections of the TUC-PT as well as the grammatical contents that are tested via the individual methods, Bachman's test method facets will not be applied to the two tests in the present section. Nevertheless, the following chapters will incorporate some of the facets just mentioned, especially those related to "nature of language input and expected response" (Bachman 1990: 130) as these are most clearly related to the linguistic/grammatical contents that are part of the focus of the test items.

5. Testing Tenses and Prepositions in the WBPT and the TUC-PT

In the following chapters, two sections of the WBPT will be compared with the grammar sections of the TUC-PT. Special emphasis will be put upon the different test methods and item formats used and the grammatical contents that are being tested. The composition and design of items in the WBPT is the primary interest of this paper. Therefore, after a brief account of the methodological approach that was used and the concepts of tense and prepositions has been given, the following analysis will start with detailed descriptions and discussions of particular tense and preposition items in the WBPT, the findings will in a second step be compared with what can be found out about some of the items in the TUC-PT focusing on these two areas of English grammar. Contrary to the formal description of the two tests (c.f. chapter 4), in the following sections, the first step of the analysis will be based on some sections of the WBPT which will in a second step be compared with the grammar sections of the TUC-PT.

---

[29]The main difference between reciprocal and non-reciprocal input-response relations consists in the fact that the former includes a certain amount of interaction between "at least two parties" which goes along with such a relation including certain communicational aspects, such as giving and receiving feedback, following a certain "communicative goal" etc. (Bachman 1990: 149). In contrast to that a non-reciprocal relationship between the input of and response to a given test task does not involve any "interaction between language users" (150). Finally, an adaptive relationship between input and expected response implies that "the input" of a test task "is influenced by the response, but without the feedback that characterizes a reciprocal relationship" (150f.).

5.1 Methodological Approach

As has been mentioned with reference to the contents of the two tests that are to be compared in this paper, both focus on a variety of grammatical topics. The present analysis will, however, be restricted to the testing of tenses and prepositions as two important areas of grammar. This restriction is, amongst other things, due to the fact that a focus on more than two grammatical categories would exceed the space provided for the present analysis. The choice of tenses and prepositions can further be justified as these two areas of English grammar can generally be considered as very frequent sources of error for broad groups of (German) learners of the language. Moreover, the selection of the two categories allows a discussion of items that focus on aspects related to VP (tenses) as well as NP structures (prepositions).

As the WBPT in its entirety presents test takers with a relatively high amount of test tasks and items, i.e. five tasks with a total number of 95 items, the present analysis will be restricted to two sections of the test, i.e. section one and section two, each of which employs a different testing technique or method. The analysis of the TUC-PT will be restricted to the grammar sections of the test, i.e. sections 2.1; 2.2 and 2.3, excluding section 1, the task focusing on spelling errors.

Following the identification of tense and preposition items in the above mentioned sections of the two tests, test taker responses to these items were analysed and documented. For the WBPT the total number of test takers that were analysed varies. This is first of all due to the fact that for section one and two of the test there exist four different versions the testees are provided with via a random principle.

The number of test takers for each version of section one and two that were included into the analysis presented here are outlined in Table 1 below.[30]

	task version	number of test takers
WBPT section one	one	43
	two	33
	three	32
	four	34
	total	142
WBPT section two	one	21
	two	26
	three	30
	four	17

---

[30] In order to have some control over the composition of the testing population, only those testees that completed the demographic questionnaire were included into the analysis.

	total	94	

(Table 1: Distribution of test takers across WBPT task versions)

It has to be pointed out that the number of test takers in the WBPT does not only vary within one and the same task but also from section to section. This can be explained by the fact that the WBPT lacks a number of characteristics of a testing situation in the strict sense of the word. Therefore, each individual doing the test has the possibility to stop the test taking process whenever (s)he intends to. Thus, the analysis showed that the target group becomes smaller with an ascending number of test sections. In contrast to that, for the TUC-PT the number of testees remains steady from section to section. The total number analysed for the purposes of this paper is 137.[31]

Within the analysis of test taker responses to items in the WBPT and the TUC-PT focusing on prepositions and tenses the percentages of test takers within the two target groups who answered the respective items incorrectly were calculated. From this a number of useful insights can be gained with respect to items in both tests which resulted in particular problems on the side of the test taker. For those sections of the two tests which employ a multiple choice method, the distribution of error rates[32] across the distractors for each of these items was determined. This practise allows an identification of the respective amount of attractivity of distrators for the target groups. For section two of the WBPT, the C-test the percentages of the respective incorrect letter combinations that were supplied for each item were calculated. These particular figures were also integrated into the description and analysis of items as they provide useful insights into the type of incorrect answers that an item evokes.[33]

Test taker responses were taken as a basis for a detailed analysis of tense and preposition items in sections one and two of the WBPT and sections 2.1; 2.2 and 2.3 of the TUC-PT. For this analysis, items which proved to be particularly problematic for the respective target groups were of special interest. However, the present paper is not primarily concerned with learner errors but with a deep discussion of the design and focus of test items. For such purposes, items which were answered incorrectly by a relatively high number of test takers are normally most interesting. Nevertheless, the following chapters will at some instances also include items that did not seem especially difficult for the test takers but which add particularly interesting aspects to the analysis,

---

[31] Test taker scores for the sections of the WBPT and the TUC-PT discussed in this paper are provided in Appendices II and III.

[32] Although generally a distinction is drawn between errors as reflections "of L2 learners' mental knowledge of the second language" (Johnson & Johnson 1998: 111) and "transcient 'mistakes' or 'lapses'" (Corder 1974 cited in ibid.), the present analysis will use the term *error* to refer to any type of incorrect answer in the WBPT and the TUC-PT.

[33] Error patterns for the items discussed in this paper are provided in Appendix VI.

particularly with regard to the comparison of the different testing techniques used in the WBPT and the TUC-PT.

The analysis of items in the two tests forms the basis for comparing the WBPT and the TUC-PT along two different criteria. First, a comparison of section one of the WBPT with sections 2.1 and 2.2 of the TUC-PT, i.e. two varieties of the multiple choice method, text-based versus sentence-based gap filling tasks, will be carried out. For this comparison, aspects of the influence of the amount of context that is provided for multiple choice test items will be particularly important. Second, section two of the WBPT, a modified C-test, will be compared with section 2.3 of the TUC-PT, a text-based multiple choice gap filling activity. Questions concerning the applicability of the two different formats for the testing of grammatical categories such as tenses and prepositions will be addressed. Apart from the two broad areas of interest just mentioned the two types of comparison will include a detailed analysis of the respective items under focus, their composition, the grammatical contents they test, and, where applicable, the choice of distractors, the role of authentic base materials, questions of potential ambiguities, etc. which might add some useful aspects alongside the general criteria the comparisons are based on.

5.2 Focus on Tenses and Prepositions – A Brief Theoretical Discussion of the Concepts

In order to have a basis for the analysis of the testing of these two areas of English grammar, it is important to discuss the basic theoretical issues related to the concepts of tense and prepositions. Therefore, the following sections will provide a brief overview of the grammatical categories of tense (and aspect) and prepositions.

5.2.1 Tense

According to Greenbaum "tense is a grammatical category referring to the location of a situation in time" (1996: 253). "From a structural point of view, English verbs are inflected for only two tenses: **present** and **past**." (Biber, Johansson, Leech, Conrad, & Finegan 1999: 453). From this, it becomes obvious that common classifications of the concept of time which differentiate between present, past, and future time, are not possible "on a grammatical level" (Quirk, Greenbaum, Leech, & Svartvik, 1985: 176). It is for that reason that, strictly speaking, only the formal categories of present tense and past tense, or past and nonpast, can be found in English (ibid.). Quirk et al. explain this by the fact that "morphologically English has no future form of the verb in addition to present and past forms" (ibid.). Therefore it becomes

possible to suggest that although there exists no future tense in English which "is realized by verb inflection" (Greenbaum et al. 1990: 47), "tense and aspect [...] relate the happening described by the verb to" future as well as to present and past time (Leech & Svartvik 1994: 65). In very general terms, "the happening described by the verb" can be a state or an event. From this, "three basic kinds of verb meaning" can be established (ibid.). These include reference to a "STATE", a "SINGLE EVENT", or a "SET OF REPEATED EVENTS", which can also be referred to as a "HABBIT" (66). To all of these verb meanings the idea of temporariness can be added with the help of the progressive aspect (66). The relationship between the categories of tense and aspect shall only be discussed briefly in this section. Generally speaking, English has "two aspect constructions [...], the perfective and the progressive" (Quirk et al. 1985: 188f.). In contrast to tense, "aspect is not deictic" (188). In other words, whereas tense is frequently described as "a systematic grammatical marking of the verb" which is used to establish a reference to time. (Klein 1994: 16), aspect does not refer to the time of the happening expressed by the verb. Mair defines aspect as "im Englischen zu Tempus und Modus hinzutretende grammatische Kategorie des finiten Verbs, durch die Information über Art und Dauer der Verbalhandlung ausgedrückt wird." (1995: 124f.). What is very important is the fact that

> aspect concerns the different perspectives which a speaker can take and express with regard to the temporal course of some event, action, process, etc.: the speaker may view it as completed, as on-going, as imminent, and possibly in other ways. This 'view' is independent of the time which the event, action, process, etc. occupies on the time axis. (Klein 1994: 16).

This definition clearly implies a strong interrelatedness of the concepts of tense and aspect and at the same time fundamental differences between the two. Questions of aspect also play a crucial role for some of the items that are discussed in this paper under the headline 'tenses'. It needs to be underscored here that although the formal structure of this paper does not make a distinction between tense and aspect, the differences between the categories that have been touched on above need to be taken into account and references to the concept of aspect will be made within the chapters focusing on tenses wherever they are necessary.

5.2.2 Prepositions

In his discussion of the preposition as a part of speech, Huddleston cites a very general definition by Curme which might also be a useful basis for the purposes of the present discussion: "'A preposition is a word that indicates a relation between the noun or the pronoun it governs and another word, which may be a verb, an adjective, or

another noun or pronoun'" (1935: 87 cited in Huddleston 1984: 91). Refining this very basic definition, Huddleston refers to a number of properties of "central members of the closed class of prepositions in English. These include the functional potential of prepositions to "take NPs as complements", the functional potential of "the phrases they head" to "have a considerable variety of functions in larger constructions[34], and their invariability, i.e. the fact that "they show no inflectional variation" (1984: 336). The varieties of functions that prepositional phrases (PPs) can fulfil within a sentence might be one of the reasons why this word class is a common problem area for many learners of EFL. Another, probably more important reason for such problems can be assumed to result from the semantic properties of prepositions. Generally speaking, it can be assumed that a high number of English prepositions have a variety of meanings, which – although often historically related in one or another way[35] – may, from a learner's point of view, seem to be relatively unrelated. It is traditionally "assumed that words constitute lexical forms that are conventionally paired with meanings, and that these form-meaning pairings are stored in the mental dictionary or lexicon" (Tyler & Evans 2003: 1). If lexical items are paired with "distinct meanings"(ibid.), or meanings that appear to a learner to be distinct, this might result in considerable problems using such items correctly. The problem of "distinct meanings" being "associated with a single form" (ibid.) is especially obvious with reference to a large group of English prepositions, and the question of how to find "a satisfactory solution to the problem of how to represent the multiple meanings associated with a single linguistic form" continues to be "a central and controversial issue for linguistic theory" (ibid.). With regard to the variety of possible meanings of English prepositions, Zelinsky-Wibbelt points out that "prepositions constitute a lexical category the elements of which are highly polysemous" (1993: 10). Generally speaking, English prepositions are not only used with meanings such as "those of SPACE and TIME" and "relationships such as INSTRUMENT and CAUSE" (Quirk et al. 1985: 673). They also can have a number of grammaticised uses. Such grammaticised prepositions have "specific syntactic roles in the language that are not determined by their meanings" (Huddleston & Pullum 2002: 647). The different syntactic as well as semantic properties of prepositions mentioned above might be one of the reasons for test items in the WBPT that focus on prepositions being a frequent source of error. Although the test focuses on a variety of grammatical

---

[34]Precisely, "in clause structure" they can function as "complement [...] or adjunct [...]; in AdjP structure" they can take the form of a "complement [...] or modifier" and "in NP structure" they can function as complements as well as modifiers. (Huddleston 1984: 336). For examples of each of these potential functions of PPs (c.f. ibid.).
[35]C.f. e.g. the historical meaning relations between "prepositions of place and time" (Lindstromberg 1997: 10).

topics, the items that deal with prepositions showed comparatively high error rates. Like many computer-based tests, the WBPT consists entirely of objective test items; the same is true for the grammar sections of the TUC-PT. Four of the five sections of the WBPT as well as all of the grammar sections in the TUC-PT employ a multiple choice test method. As has been pointed out by Bachman, the term multiple choice is a rather broad one which includes a wide variety of test method facets (c.f. 1990: 115). This also means that the multiple choice method can of course have a great number of variations. It is for that reason that in the following section the multiple choice method in general will be explained briefly. The different varieties of multiple choice will the be described in the respective sections dealing directly with the individual sections of the WBPT and the TUC-PT.

5.3 The Multiple Choice Method – General Assumptions

Davies et al. define multiple choice test tasks as consisting of items which require "the **test taker** [...] to choose the correct option [...] from several given" (1999: 124). With reference to the organisation of such items, they point out that most commonly, multiple choice items include an instruction to the test taker and a stem (typically either a phrase or sentence to be completed, or a question)" (ibid.). The correct alternative "and several distractors, usually three, then follow in random order. Stimulus material, in the form of an input text (written or spoken) may be provided [...], or the item may stand alone" (ibid.). This definition of multiple choice items and their organisation can be related to Bachman's observation that *multiple choice* is a relatively broad term used to refer to a test method that can be described according to a variety of facets, including the form and nature of the input etc. (c.f. 1990: 76f.; 116-157). One of the major reasons for developing the test method facet categories described by Bachman was to account for variations within what is broadly referred to as test methods, i.e. for the fact that test methods are not "monolithic wholes" (115). With respect to the multiple choice method, variation can for example occur with reference to the input a test person is presented with. Multiple choice items can take the form of, e.g. a matching activity, they can require test takers to "complete [...] missing information", or choose the most appropriate answer to a "direct question" (Bachman 1990: 116). Moreover, multiple choice tasks can also be of such a type that they require test takers to chose or "identify [...] incorrect part[s]" of a given utterance, paragraph etc. (Bachman 1990: 116).

One of the obvious advantages of multiple choice testing techniques is the fact that for such tasks fixed answer keys can be established. Multiple choice is thus often described as a very reliable test method. As Weir points out:

> In multiple-choice tests there is almost complete marker reliability. Candidates' marks, unlike those in subjective formats, cannot be affected by the personal judgement or idiosyncracies of the marker. The marking, as well as being reliable, is simple, more rapid and often more cost effective (1990: 43).

Since they can be scored objectively multiple choice tests are particularly useful in computer-based testing environments, especially those which aim at making the testing and scoring procedure more effective. For such purposes, answer keys need to be unambiguous and well-defined. It seems obvious that in order to yield such answer keys and, even more importantly, to be reliable measures of language ability, multiple choice test items need to be designed with special attention being paid to a number of factors. Alderson et al. mention a variety of characteristics that multiple choice test items should possess. These include the following aspects: Most importantly, multiple choice items of whatever kind must be designed in such a way that "the 'correct' answer" is "genuinely correct"; furthermore "item writers must ensure that if the answer key gives just one correct answer, then there is only one answer" (1995: 47). As far as incorrect alternatives are concerned, each of them "should be attractive to at least some of the students" (48). The number of alternatives given for each item can vary, however,

> generally it is a good idea to have at least four alternative answers, so that the chance of a student guessing an answer is only 25%, but if it is impossible to think of a third attractive wrong answer, this it is sensible to have only three alternatives for some items" (48).

This statement hints to the problem of test persons' guessing answers which needs to be accounted for when designing multiple choice test items. As explained by Davies et al. there are many multiple choice items which present only three alternatives (1999: 124), which still seems to reduce the possibility for guessing the correct answer to a justifiable extent. Dichotomous items, however, i.e. items that provide test takers with only two alternatives to chose from are therefore judged by Alderson et al. to be "generally unsatisfactory, as there is a 50% possibility of getting any item right by chance alone" (1995: 51).[36]

From what has been mentioned above, it becomes obvious that for the design of multiple choice test items, keeping the possibility for guessing as small as possible and

---

[36]Another important aspect that needs to be taken into account is the fact that "the correct alternative should not look so different form the distractors that it stands out from the rest. It should not be noticeably longer or shorter, nor be written in a different style" (49), which would again increase opportunities for guessing the correct alternative or identifying it by means other than the actual language abilities focused on by an item.

the definite identification of the correct answer(s) are major considerations. Another, very important aspect that needs to be taken into account is the fact that "multiple-choice items should be presented in context" (48) in order to be unambiguous. One of the reasons for this requirement is the fact that multiple choice items generally do not provide test takers with the possibility to explain why a given alternative was chosen, so presenting them within a suitable amount of context can help reduce ambiguity that could otherwise confuse test takers in their decision for an item and testers in scoring it. Furthermore, presenting multiple choice test items as well as any other kind of test item in context can help to make tasks more realistic because in real-life language use situations there are hardly any instances of language use which occur in isolation from a certain kind of language context.[37] Questions concerning the role of contextual features for the testing of tenses and prepositions via multiple choice methods and the potential influence of context on test ambiguity will be discussed in closer detail in within the comparison of text-based multiple choice gap filling in the WBPT and sentence-based tasks of this type in the TUC-PT which will be provided below.

As will be discussed later in this paper, presenting multiple choice items within a suitable amount of (possibly authentic) linguistic context can help to render such tasks more meaningful and might create a stronger relation to aspects of actual real-world communication. However, this can not compensate for the restrictedness of multiple choice items to language recognition skills. This restriction of test takers' response freedom can be considered to be one of the most significant disadvantages of the method. As mentioned in chapter 4.4, objective test methods such as the multiple choice technique have often been the target of criticism for focusing on receptive rather that productive skills and for not focusing on test takers' ability to actually communicate with the help of a language. With reference to the high amount of multiple choice items that can be found in computer-based language tests Alderson criticises that objective test formats such as "multiple choice questions [...] have been given a new lease of life because they are easily adapted to computer-based testing, when their validity and value have begun to be questioned in testing research"(1996: 249). He further points out that "the ease of the use of multiple choice [...] in the computer age threatens to inhibit test developers from seeking new and more valid ways of assessing language proficiency and language learning" (ibid.)

---

[37]According to Rea-Dickins "context and the function(s) that grammatical items perform within a text to create meanings, are [...] important features of communicative grammar" (1991: 120). Nevertheless, it cannot be assumed that presenting test items in context is a sufficient condition for making a test authentic or communicative; in fact "some contextualised texts may be very uncommunicative" (128).

In spite of the significant role, the multiple choice method seems to play in present-day computer-based language testing, it needs to be kept in mind that the abilities measured by most multiple choice tests are "not quite the same as the ability to produce and use the correct form [a test taker might have chosen in a multiple choice test, K.U.] in real-life situations" (Heaton 1988: 11). With regard to the disadvantages of the multiple choice method, Weir points out that "there is considerable doubt about their validity as measures of language ability. Answering multiple-choice items is an unreal task, as in real life one is rarely presented with four alternatives from which to make a choice to signal understanding" (1990:44).

Nevertheless, multiple choice test items have a number of advantages, and continue to play an important role in many testing situations but particularly those which aim at an explicit assessment of grammatical competence. Although the technique bears a number of disadvantages such as the restriction to the assessment of language recognition rather than production mentioned by Heaton (1988), it can still be considered to be a very effective and reliable means of measuring a test person's (receptive) knowledge of certain aspects of a language. The relatively significant role that multiple choice techniques continue to play in the domain of grammar testing can, amongst other things, probably be explained by this fact. As mentioned by Rea-Dickins, test items which explicitly aim at measuring the understanding of the grammatical structures that underlie a language, or as she calls it "explicit system-focused testing" of grammar still have a certain relevance in a number of testing situations (2001: 29). As has been explained in chapter 3.1, it can be claimed that one such situation could be the testing of the grammar competence of students of English and American studies.

Apart from still being a valuable test instrument in contexts like the one just mentioned, even such restricted test methods as the multiple choice technique can be adapted to a certain extent, e. g. via material's selection or possibly also via the tester's choice of distractors, to the demands of present day communicative language teaching and testing. Nevertheless, it needs to be emphasised that it is more or less impossible to develop a valid test of the ability to appropriately use the grammatical structures of a language or even overall language proficiency which consists entirely of multiple choice items. With reference to the potential of multiple choice test items to be a measure of communicative grammatical ability, Rea-Dickins points out that "response format such gap filling and multiple-choice (whether it is for grammar testing or other language skills) can never be considered communicative" (1991: 129).

5.4 Multiple Choice Grammar Testing in the WBPT and the TUC-PT

Applying the restriction of the multiple choice method to language recognition skills and their relative inappropriateness for the testing of communicative language abilities to the case of the WBPT and the TUC-PT it needs to be clearly underscored that although the grammar sections of the TUC-PT are entirely multiple choice, they need to be viewed within the context of the whole test which consists of several components focusing on different language skills and abilities at many instances in combination with one another. Grammar is therefore not only tested explicitly in the grammar sections but also implicitly within the oral interview or the writing section of the test. As the WBPT is designed in order to be a practise opportunity for the grammar sections of the TUC-PT, it cannot claim to be a measure of the overall ability of test persons to communicate in English or, in general terms, their general English language proficiency. Nevertheless, the WBPT reflects the particularly high importance that the TUC-PT attaches to explicit grammatical knowledge.[38]

Given the relative restriction of test takers' creativity of language use that is characteristic for the multiple choice method, the WBPT tries to introduce a certain extent of variation within its different multiple choice sub-sections. This is done for instance via the selection of material's from different, often authentic, sources or the different linguistic contents the test items focus on. These do not only include traditional grammar topics, they also try to incorporate at some aspects such as text cohesion and coherence, the organisation of written as opposed to spoken discourse etc. Such variations within grammar test tasks have been observed by Rea-Dickins to be characteristic of present-day changes in the explicit testing of grammar (2001: 28). Nevertheless, the WBPT is first of all a test of those areas of English grammar that are also part of the TUC-PT and thus needs to be viewed as an explicit instrument for the assessment of grammatical competence as just one aspect of general language proficiency.

In the following, items in section one of the WBPT which focus on English tenses and prepositions will be analysed. The findings will then form the basis for a comparison with sections 2.1 and 2.2 of the TUC-PT. Section one of the WBPT is a text-based multiple choice task, sections 2.1 and 2.2 of the TUC-PT measure grammatical competence on the basis of the sentence or very short sequences of sentences. Accordingly, the comparison will allow valid inferences about the role of

---

[38]This strong focus on explicit grammatical competence is also reflected in section four of the TUC-PT and section four of the WBPT which focus directly on test takers' competence in the terminology of English grammar, i.e. testees are not only required to have a detailed knowledge of the formation and use of grammatical structures but also of the metalanguage used to denominate such structures.

context for the testing of tenses and prepositions with the help of the multiple choice method.

The significant role that the context seems to play for multiple choice test tasks has for instance been acknowledged by Alderson et al. (1995: 48). It is also reflected in Bachman's categorization of test method facets. According to him the "degree of contextualization" is one of the major aspects of the facets of test input and expected test response (1990: 131).[39] On the basis of Cummins (1983), he explains "the notion of 'context embeddedness'" (cited in Bachman 1990: 131) of language which includes "a wide range of meaningful linguistic, paralinguistic, and situational cues in the context" (ibid.). This characterization of the context-embeddedness of language does not only refer to the immediate linguistic context that surrounds a respective item but also to features such as the amount of familiar topical information that is presented by or involved in finding the correct solution to an item and a number of other factors. As for multiple choice items test taker responses are in most cases fairly restricted with regard to the degree of contextualization, it is the input that such items provide for which the context plays an extremely important role. As pointed out by Bachman "the more context-embedded the input is, the more likely the test taker will be able to respond to its propositional content." (132). As has been claimed by Alderson et al. (1995: 48), it is thus essential for a test method which is relatively restricted as far as the test takers' response freedom is concerned, to provide enough context in the input in order to make the item understandable and unambiguous. As shall be explained later on, the context a multiple choice grammar item is presented in assumes a quite special role as the choice of a particular grammatical pattern, e.g. a certain tense form, is in most cases highly dependent on the respective context. The amount of context that is needed for making multiple choice items dealing with tenses and prepositions will be one of the major considerations of the following part of this paper.[40]

Although the influence of the amount of context provided constitutes the major focus of the following sections, aspects such as the need for genuinely correct solutions and incorrect distractors and the discussion of individual grammar problems will constitute an integral part of the analysis as well.

---

[39]Bachman makes this assumption not only with reference to multiple choice but to a variety of test tasks in general.

[40] Apart from the facts just mentioned there exist other interesting aspects that are related to the influence of contextual features on the choice of a particular item in multiple choice test tasks. One of them is the fact that presenting multiple choice items in isolation from the context, e.g. within short sentences or phrases which do not present enough linguistic and situational context in order for test takers to make a definite decision for one of the alternatives s/he can chose from, might not only render the item ambiguous but also decrease test fairness. The latter is due to the fact that such items can, under certain circumstances, easily confuse test takers in such a way that an incorrect alternative is chosen which would otherwise, i.e. in a testing situation which presents enough contextual information, not have been chosen.

5.4.1 Multiple Choice Gap Filling On Text Basis – WBPT Section One

Section one of the WBPT is a gap filling activity requiring testees to complete a text which contains 25 gaps. For each gap there are three alternatives given from which the test taker has to choose the most appropriate one.[41] Examples for the different versions of section one are provided in Appendix I. The importance of presenting multiple choice items within an appropriate amount of context has already been discussed in the previous chapters. In general, gap filling tasks which are based on a complete text dealing with information that can be assumed to be familiar to most test takers provide a relatively high amount of context-embeddedness in the sense of Bachman.

Like the multiple choice test method in general, gap filling tasks may exist in a number of different varieties. The individual items in traditional gap filling tasks are derived from a deliberate selection of gaps which is based on the respective contents a given test is supposed to measure. Such tasks are a very common measures "of reading or listening comprehension, or of grammar" (Davies et al. 1999: 66). In order to guarantee an appropriate amount of test fairness as well as the context necessary for successful completion of items, the texts which are the basis for section one of the WBPT focus on topics that every test taker can be assumed to have a certain knowledge of. There exist four versions of section one and every test person who logs on to the WBPT is supplied with one of these versions via a random principle. Text one deals with the United Kingdom, certain general problems arising out the regionalism typical for the country and raises the question whether there exists one British national identity (c.f. Appendix V, "British National Identity I"). The text was taken from *Britain in Close Up*, a textbook dealing with a variety of topics concerning the British culture, politics, religion etc.[42] The topic of text two, which was taken from the BBC News webpage (c.f. Appendix V, "Brenda Fassie I"), is the life and professional career of African popular singer Brenda Fassie. Text three, which is again an authentic news text taken from the web edition of CNN (c.f. Appendix I, "New Teacher Standards I") deals with the American educational system. Like text two, text four also originates from the BBC News WWW edition and centres around a prehistoric site which has been rediscovered by archaeologists in India (c.f. Appendix I, "Ancient Indians I").

---

[41] A task of the same type can be found in the TUC-PT; section one of the WBPT can, thus, be described to be the only component of the test which directly adopts one of the varieties of the multiple choice method used in the TUC-PT.

[42] C.f. McDowall (1999).

As can be seen from this description, all of the texts focus on topics that do not call for a high amount of specialised knowledge on the side of the test taker. The texts were at some instances shortened but only slightly adapted and can therefore be assumed to include a relatively high amount of authenticity. This is especially the case for the news texts (texts two to four). Although text one is taken from a textbook which was obviously written for learners of English, as can be seen from the vocabulary glossary at the end of the book (c.f. McDowall 1999: 204f.), it can still be assumed to be comparable to the other texts in section one as the book employs a very advanced type of English, which can at many instances be compared to that of authentic non-academic articles dealing with the same topic, and is even recommended as a textbook to students of English and American Studies at Chemnitz University of Technology (c.f. *British and American Studies – Veranstaltungsangebot Wintersemester 2004/2005*)

Although the base materials for section one were selected with a special view to equality and comparability with regard to the relative familiarity of content and the difficulty level across the four texts, the mean scores of the four texts reveal that there seem to exist differences between the four texts as far as their difficulty is concerned. For text one, the mean score is 78%, test takers who worked on text two achieved a mean score of 69%. The figures for texts three and four are 61% and 76%, respectively. From this, it can be seen that the mean score was highest for the text taken from the learner textbook. The fact that scores were lower for the news texts can surely be explained to a certain extent by the particularities of the language of the news which generally tend to pose a high amount of problems to learners of English. However, it is not possible to conclude from this that all the news texts that are part of section one can be treated as equal with regard to their difficulty level. As can be seen from the mean scores, there is considerable variation within the mean scores obtained for texts two to four. This can on the one hand be due to changing difficulty levels within the input as such and on the other hand to the selection of the items themselves. In the following, some of the items in section one focusing on tenses and prepositions will be examined in closer detail.

5.4.1.1 Testing Tenses

Taken together, all the four texts in section one contain 15 items focusing on tenses. These are more or less evenly distributed among the four texts. Text one, three and four contain four items, measuring competence in the use of tenses, respectively. Text two includes three such items. As mentioned earlier on, the text material for section one was only slightly adapted, the materials were carefully selected with

reference to their topics, their degree of authenticity and their language difficulty level. However, in order not to make the texts artificial documents whose sole purpose is to serve as material for the testing of as many grammar points as possible, they were not changed with reference to their grammatical structure, i.e. items in the texts were chosen on the basis of the original version of the respective text. This way of keeping as much of a texts natural composition and (grammatical) structure leads to certain restrictions as far as the variety of grammar topics that can be tested on the basis of a text is concerned. As all of the four texts in question bear a certain relationship to past events which can be described as a result of their topics which all have a certain connection to past events and developments. Accordingly, a relatively high number of the items in section one that focus on tenses can be said to measure competence in the use of past tenses. Via the selection of distractors, it is, however, at many instances possible to measure such ability in relation to the use of other tenses.

One example for an item that tests the use of a present perfect tense as opposed to present tenses is item one of version one:

[5.4.1.1.I]

The United Kingdom of Great Britain and Northern Ireland, to give it its formal title, is a highly centralised and unitary state. Its largest part, England

_____

1) has played[43]
2) is playing
3) plays

a leading role in the country for almost 1,000 years. [44]

From the error rates for [5.4.1.1.I] it can be assumed that the distractors for this item were attractive to a relatively high number of test takers. 37% of the 43 subjects who were provided with text one in section one of the WBPT decided for one of the incorrect alternatives. Alternative two, i.e. the present progressive form of *to play* was chosen by 14% of the total test population, for alternative three the figure is 23%.

---

[43]Throughout this paper, the respective correct or most appropriate answers to multiple choice items are underlined.
[44]Throughout the following sections, items will be presented within a clause, or one or more sentences in order to provide an appropriate amount of context.

Looking at alternatives two and three, it becomes obvious that the item deals with the opposition of the present perfect which can be used with a certain reference to the present and present tenses (either with simple or progressive aspect) as such. Comparing the present perfect to the simple past, Quirk et al. observe that "the present perfective differs from the simple past in relating a past event/state to a present orientation" (1985: 190). With the choice of alternative one for [5.4.1.1.I], test takers prove that they are able to realize the relation that is expressed with the help of the present perfect tense between a past state that continues into the present. Such a type of use of the present perfect is frequently labelled "state present perfect" (Greenbaum 1996: 270). It is one of the common characteristics of the state present perfect in particular as well as the present perfect in general that utterances made in this tense frequently "contain an expression denoting a period of time extending from some time in the past to the present" (ibid.). In the case of [5.4.1.1.I] the reference to such a time period is made with the help of the preposition *for* in combination with a noun phrase signifying the length of this period. It is this fact that very clearly determines alternative one as the correct solution to item [5.4.1.1.I]. The characteristic of signifying "past time 'with current relevance'" that is frequently associated with the present perfect (Quirk et al. 1985: 190) is most definitely expressed by the use of *for* in combination with a particular duration, in this case *almost 1,000 years*. Text one describes the history of the UK as well as a number of general facts concerning the country's structure and problems. Accordingly, the basic tense of many of the passages of the text is the simple present and it can be concluded from this that the time of utterance (TU) for [5.4.1.1.I] is in the present. The leading role of England within the country can therefore be assumed to have started almost 1,000 years ago and to continue up to the time of speaking, i.e. the present, and even into the future. From this it can be assumed that the context of the sentence surrounding the item as well as the context of large parts of the base text have a strong influence on the choice of alternative one as the most appropriate solution to [5.4.1.1.I].

The two distractors for [5.4.1.1.I] are both present forms of the verb *to play*. Interestingly, the duration of particular events and situations that started in the past and continue up to the present or the TU which is generally expressed with the help of a verb form in the present perfect or present perfect progressive in English, is frequently denoted by the *Präsens* in German. Therefore, it could be hypothesized that the item does not only focus on the use of different tenses and aspects within the English language, it also places a certain emphasis on language interference problems that might arise out of the characteristics of a test taker's mother tongue. Therefore, the choice of

alternative three and especially alternative two can, under certain circumstances, be described as a result of negative transfer of German tense usage into the English language. However, the present paper is not primarily concerned with German-English language interference problems and, although a very high proportion of the target group has German as their mother tongue, it cannot be claimed without any further investigation that a decision for alternatives two or three is due to the influence of the test takers' mother tongue. What can, however, clearly be maintained at this point is the fact that, in contrast to German, past events that bear a relation to the present are not normally expressed by a present tense in English which makes it possible to clearly identify alternatives two and three as incorrect.

According to Quirk et al., there exist three general uses of the simple present which include the so-called state present which is "used without reference to specific time" (1985: 179), the habitual present which like the state present implies "an inherently unrestricted time span" and expresses "a whole sequence of events repeated over the period in question" (ibid.), and the instantaneous present which occurs "where the verb refers to a single action begun and completed approximately at the moment of speech" (180). Comparing these possible uses of the present tense with the situation provided by [5.4.1.1.I], none of them applies to the example.

As mentioned earlier on, alternatives two and three were basically chosen in order to measure test takers' understanding of the present perfect as a tense that expresses a certain relationship between the past and the present as opposed to present tenses which cannot be used with this meaning in English. The fact that the state referred to in [5.4.1.1.I] has been going on for a longer period of time, might have led some test takers to chose alternative two, the present progressive form of the verb *to play*. This can be explained by the fact that the progressive aspect generally "indicates a happening in progress at a given time" (Quirk et al. 1985: 197). The progressive aspect can have three components of meaning, "not all of which need be present in a given instance" (198). These are the duration, the limited duration, and/or the possible incompleteness of a happening (ibid.). Both the first and the third of these meaning components can quite easily be associated with the situation provided by [5.4.1.1.I]. However, like alternative three, alternative two only bears a direct relationship to the present as opposed to the past time reference included in the situation being made.

Like [5.4.1.1.I], item four of version two, puts a certain emphasis on competence in the use of the present perfect tense. In contrast to [5.4.1.1.I], however, the item focuses on the contrast between the present perfect and the simple past as well as the past perfect.

```
[5.4.1.1.II]
Brenda _____


1) had been named
2) has been named
3) was named


after US country singer Brenda Lee
```

All of the alternatives to the item are presented in passive voice which might add to the seemingly high difficulty of the item for many test takers. 39% of a total of 33 test persons, chose one of the distractors, 12% and 27% of all test persons chose alternative one or two, respectively. Nevertheless, the main focus of the item is clearly on the use of tense and aspect as all of the alternatives are presented in passive voice and the distinction between active and passive can therefore not be considered to be part of the grammatical contents measured. The correct solution to [5.4.1.1.II] is alternative three, i.e. the simple past of the verb *to name*. The correctness of this alternative can easily be proven with the help of the descriptions of the possible meanings of the past tense in English provided by Quirk et al. According to them, the past tense contains a combination of the following "two features of meaning" (183):

> (a) The event/state must have taken place in the past, with a gap between its completion and the present moment.
> (b) The speaker or writer must have in mind a definite time at which the event/state took place (ibid.).

Such definite time reference is usually expressed with the help of "past time-position adverbials, such as *last week*, *in 1932*, *several weeks ago*, *yesterday*" and so on (ibid.). Feature (a) can without any doubt be applied to the situation provided by [5.4.1.1.II]. The most obvious prove for this can already be found in the first sentence of the text: "Brenda Fassie – whom Time magazine called the "Madonna of the Townships" - has died in South Africa […]" ("Brenda Fassie II", c.f. Appendix I). Moreover, the whole of

text two describes various aspects of Fassie's life and career, which clearly indicates that there exists a certain time gap between the time of the naming of the person in question and the TU. Therefore, should the context of the sentence surrounding the item not be sufficient for test takers to recognize this fact, the context provided by the whole of the text can serve as a supporting cue for deciding on the correct alternative. This relatively simple example hints to the importance of the context for successful completion of multiple choice items claimed by Alderson et al. (1995: 48) and to the importance of contextual factors for language tests in general discussed by Bachman (1990: 131ff.).

Feature (b) which is described by Quirk et al. to be typical for the past tense is not explicitly identifiable in [5.4.1.1.II], however, alternative three can definitely be determined as the correct solution to the item, as "it is not necessary for the past tense to be accompanied by an overt indicator of time" (Quirk et al. 1985: 184). The reference to a specific point of time in the past can also be derived from a number of other sources, these include: (a) the immediate or local situation; (b) the larger situation of general knowledge; (c) what has been said earlier in the same sentence or text; or (d) what comes later on in the same sentence or text" (ibid.). Applying this to [5.4.1.1.II], it becomes obvious that it is especially sources (b), (c) and (d) which are of relevance for the item. The previously mentioned importance of the contextual information provided by the text the item is part of becomes especially obvious with reference to sources (c) and (d). Quirk et al. identify three possible meanings of the past tense in English, these are event past, state past and habitual past (1985: 186). The meaning which is most appropriate with reference to example [5.4.1.1.II] is that of the most common one of those cited above, the event past. This type of past tense "refers to single definite event in the past" (ibid.), in the case of [5.4.1.1.II], the naming of Brenda Fassie.

Taking all the possible characteristics of the past tense mentioned above into account, it seems that the item very explicitly calls for the choice of alternative three as the correct solution. Nevertheless, a relatively high proportion of test takers decided for one of the two distractors. Alternative two was more attractive to the subjects who answered the item incorrectly (27% of the total test population). This might probably be due to certain overlaps that exist between the meanings of the simple past and the present perfect. Like the simple past, the present perfect can refer to past events. A crucial difference between the two tenses consists, however, in the fact that the present perfect usually refers to "indefinite event(s) in a period leading up to the present" (Quirk et al. 1985: 192), whereas to justify the use of a simple past tense the event in question must be completed in the past (183). Clearly, alternative two focuses on test takers' understanding of this difference which is one of the main differences between the

simple past and the present perfect.

Though not as frequently chosen as alternative two, alternative one, the past perfect form of the verb *to name* was attractive to almost one third of the test takers who answered the item incorrectly (12% of the total test taker population). According to Greenbaum et al., the main distinguishing characteristic of the past perfect tense is the fact that it "refers to a time earlier than another past time" (1990: 53). Thus, if [5.4.1.1.II] would call for a completion with the help of a past perfect tense, the context surrounding the item would have to contain a reference to a past event or state which could be identified as following the naming of Fassie.

[5.4.1.1.II] is followed by a clause containing a past form:

[5.4.1.1.III]

, and began to perform _____ (at, with, --) five years old.

The past form contained in this clause can, however, not be described as referring to a past event following the past event that is part of [5.4.1.1.II]. The relation between the two clauses does not place a primary emphasis on the naming of Fassie preceding the time of her beginning to perform. Rather, the fact that the two clauses are joined by the conjunction *and*, implies that the primary relation between the two is one of coordination rather than a temporal one where the action expressed by the verb of one clause precedes that expressed by the verb of the other. *And* is very commonly used for "the usual kind of coordination", i.e. simple coordination (Greenbaum et al. 1990: 271). In the case of simple coordination, "a single clause or clause constituent is linked to others that are parallel in meaning, in function, and (generally) in form" (ibid.). Therefore the relation between the verb form in [5.4.1.1.II] and that in [5.4.1.1.III] can be described as a parallel one, or in other words, a series of events or actions in the past. Apart from being a parallel one, the relation between the two clauses that are part of the sentence surrounding [5.4.1.1.II] is not one of coordination of two "complete independent clauses" (272). Rather the sentence contains a "coordination of predicates", with one and the same shared subject (c.f. ibid.). The fact that the context of [5.4.1.1.II] does not put special emphasis on a certain action, state or event following the naming of the subject, i.e. Fassie, clearly implies that alternative one is an incorrect solution to the item.

Summarizing, the analysis of [5.4.1.1.II] showed that contextual information mentioned at several instances within the text have a crucial influence on the tense/aspect to be identified as most appropriate. In order to identify alternative three as

the appropriate solution to [5.4.1.1.II], the way the clauses immediately surrounding the item are joined as well as several pieces of information originating from the broader context of text two can serve as important cues. In the case of an identification of the past perfect verb form as incorrect, it can be assumed that the clause immediately following the item and the way it is combined with [5.4.1.1.II] has a relatively important role to play.

For a number of verbs in section one of the WBPT, the way clauses are joined plays a certain role with regard to the selection of the appropriate tense form by the test taker. Items can, amongst other factors, often be inferred from the way clauses are joined and the type of the conjunctions used for this purpose. A further example for this is item 15 of version three. This item will be presented in combination with the item directly preceding it as in this case a strong relation between the item in question and the context directly preceding it can be noted.

<div style="border:1px solid">

[5.4.1.1.IV]

The result is that

1) during

2) when

3) <u>while</u>

400 districts in Nebraska, 375 districts in Montana and 80 districts in Oregon qualified for the extra time,

</div>

<div style="border:1px solid">

[5.4.1.1.V]

no districts in South Carolina or Alabama _____

1) had qualified.

2) <u>qualified.</u>

3) qualify.

</div>

In contrast to what has been said with reference to [5.4.1.1.II], the relation between the clauses constituting the context for the items in [5.4.1.1.IV] and [5.4.1.1.V] is not one of coordination but subordination.[45] [5.4.1.1.IV] calls for a conjunction which

---

[45]Clearly, [5.4.1.1.IV] focuses on the use conjunctions as opposed to prepositions. Conjunctions and prepositions have certain features in common; both have for instance, "a relating or connecting function" (Quirk et al. 1985: 659). One of the crucial features that distinguishes prepositions from conjunctions,

entails concessive/contrastive meaning aspects in combination with a certain time reference. What is very important with reference to a test taker decision for the correct verb tense in [5.4.1.1.IV] is the fact that the correct solution to [5.4.1.1.IV] must entail semantic aspects of a certain temporal parallelism, i.e. the fact that the two situations expressed in the clauses joined by the respective conjunction were taking place during the same period of time.[46] Having recognized this fact, the decision for the correct alternative in [5.4.1.1.IV], i.e. the simple past form of the verb *to qualify* should become a relatively straightforward one. As mentioned earlier on, the correct alternative for [5.4.1.1.IV] implies a fairly high amount of temporal parallelism. Accordingly, the relation between the two situations expressed in [5.4.1.1.IV] and [5.4.1.1.V] can be assumed to reflect this parallelism. Accordingly, the tense of the verb in clause one must also be the tense of the verb in clause two. The correct selection of the simple past tense for [5.4.1.1.V] can therefore be viewed as being influenced to a certain extent by the selection of the correct conjunction in [5.4.1.1.IV]. Nevertheless, even if a test taker incorrectly chose alternative two, *when*, for [5.4.1.1.IV], the correct choice of the simple past tense in [4.1.1.1.V] would not automatically become impossible as the primary meaning of the conjunction *when* also entails a certain reference to parallel temporal structures as well as aspects of concessiveness (c.f. *Oxford English Dictionary (OED)*: s.v.). Accordingly, the relation between [5.4.1.1.IV] and [5.4.1.1.V] shows that the choice of the correct alternative to an item is to a certain extent influenced by the choice of a respective alternative to another item. Nevertheless, the fact that all the three alternatives provided in [5.4.1.1.IV] entail a certain amount of temporal parallelism, also shows that one inappropriate choice does not automatically have to lead to a series of inappropriate choices which would in turn decrease test fairness to a considerable extent.

Comparing the number of test takers who answered [5.4.1.1.IV] incorrectly with the proportion of the test population who selected an incorrect alternative to [5.4.1.1.V] shows that there might exist a certain influence of a respective choice for [5.4.1.1.IV] on finding the correct solution to [5.4.1.1.V]. It can nevertheless be claimed that there exists a certain independence between the items. This can be illustrated by the error rates for the two items: 28% percent of the total number of test takers selected a wrong

---

however, is the fact that the former introduce complements which are nominal or nominalized", (Quirk et al. 1985: 660), i.e. NPs. In contrast to that, the latter normally occur in combination with clauses which makes it possible to identify alternative one, the preposition *during*, as incorrect. The distinction between alternatives two and three, however, is not as easy to explain. Both conjunctions can imply temporal as well as concessive aspects of meaning and the difference between the two seems to be a relatively complex one (c.f. *OED*: s.v.).

[46]This seems to be most explicitly expressed with the help of *while* which has the primary conjunctive use of "during the time that" (*OED*: s.v.).

alternative for [5.4.1.1.IV]. Comparing this to [5.4.1.1.V], it becomes obvious that the error rate is much higher for the item as 56% of all test takers made an incorrect selection here. It has furthermore to be maintained that the distribution of the selection of alternatives for [5.4.1.1.V] is rather unbalanced. While 53% of the total number of subjects chose alternative one, the past perfect form of the verb *to qualify*, only 3%, or one out of 32 test takers chose alternative three, the simple present form of the verb. This implies that alternative three is not a very useful distractor for [5.4.1.1.V]. It can be assumed that the majority of test takers was able to recognize the fact that the construction in question calls for a certain past form of the verb. However, more than half of the total number of test persons were not able to realize that the verbs in the two clauses discussed above express, amongst other things, a parallel temporal relation which makes it impossible to use a past perfect form of the verb. As mentioned by Quirk et al., a past perfect verb tense "usually has the meaning of 'past-in-the-past', and can be regarded as an anterior version either of the present perfective or of the simple past" (1985: 195). For the item in question no such relation between the two clauses in question seems to be remarkable.

The item discussed above clearly shows that contextual features of the sentence surrounding an item such as the joining of clauses via particular conjunctions and the semantic aspects that are expressed by them as well as the tenses of verb forms contained in such sentences can have a particularly strong influence on the identification of a particular verb tense as most appropriate solutions to test items. It can thus be claimed that for [5.4.1.1.V] the context of the sentence surrounding the item plays the most role for the identification of alternative two as most appropriate or correct.

Apart from putting a certain emphasis on the type of conjunctions used for joining clauses, section one of the WBPT also contains a certain number of items focusing on the use of tenses in reported speech constructions. This can be explained by the fact that especially texts, two and three, i.e. the news texts, tend to report ideas and opinions of several persons in an indirect way. One example for an item focusing on the use of tenses within reported speech is item 22 of version three.

[5.4.1.1.VI]

But in a recent interview, Ray Simon, the assistant secretary for secondary and elementary education, said the criteria _____

1) is

2) was

3) <u>were</u>

being re-examined.

Taken together, texts two to four of section one of the WBPT contain four items focusing on tense usage in reported speech constructions. The error rates for all of these items were very high, 50% or higher. [5.4.1.1.VI], however, can be identified to have yielded the highest error rate when compared to all other items just mentioned. 84% of a total of 32 test takers answered the item incorrectly. [5.4.1.1.VI] is a particularly interesting item not only because of its high error rate but also because it can be assumed to focus on a combination of more than one aspects of grammar or grammatical and/or textual competence in the sense of Bachman (c.f. 1990: 87) which might be one of the reasons for the seemingly high difficulty rate of the item. With reference to the use of tenses in reported speech, the item focuses on the contrast between the simple present and the simple past, i.e. aspects related to backshifting and the sequence of tenses in the reporting and the reported clause. Apart from this, the contrast between singular and plural uses of the verb *to be* is in the centre of attention. Accordingly, the item can be assumed to focus on a combination of competence in the use of tenses and the use of singular and plural forms of the verb.

Whether a verb is used in the singular or in the plural, obviously depends on the number of the subject of a respective sentence or clause. In the case of the reported clause in [5.4.1.1.VI], the subject is the noun *criteria*, the irregular plural form of *criterion*. In English, the following six "types of plurals" can be distinguished, "regular *-s* plurals, -s plurals accompanied by modification of the base [...], base plurals [...], plurals with vowel change [...], a small set with the suffix *-en*, and foreign plurals of various kinds." (Huddleston & Pullum 2002: 1585). Clearly, *criteria* is a foreign, i.e. a Greek plural (1592). What might add to the difficulty of [5.4.1.1.VI] is the fact that there exist a relatively high number of foreign plurals in English and that all the different languages the respective words or plural forms originate from show their own distinctive patterns as far as the formation of plural nouns is concerned (c.f. 1590-1594). For plurals originating from Greek, certain patterns for words which end in *on* in the singular form can be established (1593). The most general of them is the fact that for

such words, "the foreign plural replaces *on* by *a*" (1593). Nevertheless, there exist a number of variations of this pattern. For certain words, *on* can for instance be replaced by *a* as well as by *s* (ibid.), which leads to the formation of a regular -*s* plural. Interestingly, criterion normally forms the plural along the general pattern by a replacement of *on* by *a* (ibid.).

Alternative two for [5.4.1.1.VI], although in the correct tense if the rules of backshift are strictly adhered to, can therefore be identified as an incorrect alternative, as the subject of the reported clause calls for a plural form of the verb. The analysis of the error rates for [5.4.1.1.VI] showed that a relatively high number of test takers seemed to have problems identifying *criteria* as a plural noun. This can be underscored by the fact that 56% of a total of 32 test takers decided for alternative two, the singular simple past form of *to be*, and 28% selected alternative one, the singular simple present form of the verb. As both distractors to the item are presented in a singular form, it can be assumed that all the test takers who selected a wrong alternative to the item did not realize the need for a plural form of the verb.

As far as the decision between the simple present and the simple past tense in the reported clause is concerned, 28% of the testing population decided for the simple present which as such could be judged incorrect if one applies backshifting rules in a very strict sense. According to these rules, "the past tense in the reporting verb tends to make the verb of the subordinate clause past as well" (Quirk et al. 1985: 187f.). Nevertheless, it needs to be maintained that backshift can be optional, except for situations "when the time reference of the original utterance [...] no longer applies at the time that the utterance [...] is reported" (1026). In the case of [5.4.1.1.VI], the correct alternative *were* was taken directly from the original version of the text. If one takes into account the common patterns of the sequence of tenses, the original utterance made by Ray Simon, which was then reported in the news text must have been made either in the present or in the past tense, as only these two tense forms can (in combination with simple or progressive aspects) can be transformed into a past tense in a reported clause which is introduced by a reporting clause in the past (c.f. Quirk et al. 1985: 1026).[47] Interestingly, the verb form *were being examined* in the reported utterance implies that this original utterance must have employed a progressive aspect, indicating that the process of re-examination of the criteria was in progress at the moment of speaking. Furthermore, this process does not seem to be completed at the moment of speech and

---

[47]As the original utterance made by Simon cannot doubtlessly be reconstructed here, the following discussion will be based on the assumption that the original utterance was identical in content and choice of words with the reported utterance.

reporting as the whole text deals with the problems arising out of the criteria in question: The sentence preceding [5.4.1.1.VI] for instance says

[5.4.1.1.VII]

If we said anyone could get more time, this (21) <u>would be</u> changing things dramatically.

The type of this conditional clause implies that the re-examination of the criteria being the cause for the problems described in the text might lead to an improvement of the situation in the future, but that the process re-examination which is reported was not been finished at the moment of reporting. It can therefore be expected that the tense of the original statement was not past but present, precisely the present progressive. Therefore, as the original statement seems to have referred to a process which was not finished at the time the statement was reported, backshift of tenses could be regarded as optional in this particular case. Nevertheless the decision for a past or a present tense in reported speech constructions where backshift is optional is also determined by a number of pragmatic factors. Huddleston points out that "the speaker's [or writers; K.U.] attitude to the proposition expressed in the subordinate clause" influences the decision for or against backshift. (1984: 152). If there is a certain degree of doubt concerning the truth of the propositional content of the reported clause, it would be appropriate to use backshift, whereas if the speaker or writer is absolutely convinced about the truth of the proposition, a decision against backshift would be more suitable (ibid.). As the whole of text three very critically deals with the educational policy which it describes as well as the members of the government who are responsible for it (c.f. "New Teacher Standards II", Appendix I), it can be assumed that there is a certain degree of doubt implied in the reported clause in [5.4.1.1.VI], which might also have been the original motivation for using a past tense when reporting the statements made by governmental representative Simon. From this it follows that, taking into account factors such as the pragmatic aspects just mentioned, a past tense in the reported clause would have been the most appropriate, though not the only possible selection for item [5.4.1.1.VI], even if more than one or even all of the alternatives had been presented in the plural. The fact that all the distractors are in the singular, however, relatively explicitly calls for a selection of alternative three as the most appropriate and at the same time only possible answer.

Although [5.4.1.1.VI] has been discussed under the headline tenses, it needs to be maintained that the item does not only, and possibly not even primarily, focus on the

use of tenses. As has been outlined in the previous paragraphs, the use of a past or present tense is dependent upon a number of factors some of which can be described to be to a certain extent dependent on the individual opinions and preferences of different test takers. Such factors can, however, be described to add a component to the testing situation which brings a relatively inflexible testing method such as the multiple choice technique closer to questions concerning the use of grammatical structures in real-life communication. One such question would concern the decision for a respective tense partly on the basis of pragmatic aspects. It has nevertheless, to be kept in mind that the multiple choice method, especially if to be scored via the computer, is not the best technique for testing pragmatic competencies etc. as these are usually dependent upon individual interpretations and opinions concerning the situation presented. Therefore, the choice of distractors for the item, which led to a combination of several grammatical aspects being measured, was used as a means for making the appropriateness of the correct solution more unambiguous. Much more importantly, such a combination also has a certain communicative value, as it is a crucial feature of natural communication that different aspects of grammar need to be appropriately used in combination with one another. Of course, this does not make a multiple choice grammar test a reflection of natural real-life communicative situations. However, it can, although to a relatively restricted extent, be viewed as an attempt to achieve somewhat higher degree of integrativeness and orientation towards the demands of successful real-life communication.

With regard to the significance of contextual features for the identification of the most appropriate solution for [5.4.1.1.VI] it can be summarized that features of the immediate context, i.e. the sentence the item is part of as well as the broader context of the whole of text three have a crucial role to play. The decision whether to select a singular or a plural form of the verb is motivated by the sentence, particularly the sentence subject. Nevertheless, it can be claimed that the whole text and the way particular problems are described and attitudes are expressed has at least a supporting function as it makes the need for backshift in the reported clause more or less explicit.

All the examples discussed above focused on the use of tenses with reference to main or, in the case of *be*, primary verbs. In contrast to that, item twenty of version four focuses on the use of modal verbs, in particular the use of modals with reference to past as opposed to present events and situations:

[5.4.1.1.VIII]

She believes that the people who made the motifs and (19) <u>those</u> who went to see them _____

1) must be

2) <u>must have been</u>

3) will be

physically fit and agile.

All the alternatives to this items contain modal verbs, either expressing a past or a present reference. One of the crucial characteristics of modal verbs, such as *must* and *will* is the fact that their semantics include "such concepts as volition, probability and obligation" (Quirk et al. 1985: 120). In a very general sense, "modality may be defined as" a way of influencing "the meaning of a clause" in order to "reflect the speaker's judgement of the likelihood of the proposition [...] being true" (219). A crucial characteristic of modals which is to some extent focused on by [5.4.1.1.VIII] is the fact that the meaning of modal verbs "can be divided into two types", which are reflected in the concepts of intrinsic or deontic and extrinsic or epistemic modality. Intrinsic or deontic modality "involves some kind of human control over events"(ibid.). In contrast to that, concepts of extrinsic or epistemic modality "typically involve human judgement of what is or is not likely to happen" (ibid.). In other words, "epistemic modality in language [...] usually, perhaps always, [...] relates to an inference by the speaker and is not simply concerned with objective verifiability in the light of knowledge (Palmer 1990: 7).

Both, the meanings of *must* and *will* can be described with the help of the distinction between intrinsic/deontic and extrinsic/epistemic meaning. Intrinsic *will* can generally be described to have the meaning of volition which can be further subdivided into three types depending on the force of the respective notion; the three types are willingness, intention, and insistence (Quirk et al. 1985: 229). In an extrinsic sense, *will* includes semantic components of prediction; within this category, "three related uses of *will*", future, present, and habitual senses of prediction "are to be distinguished" (228).

The basic intrinsic sense of *must* is obligation, i.e. "the speaker is advocating a certain form of behaviour" (225). In its extrinsic meaning, *must* expresses "'logical'" or "'epistemic necessity'", showing "that the speaker judges the proposition expressed by the clause to be necessarily true, or at least to have a high likelihood of being true" (224f.).

Applying all of the possible senses of the modal verb *must* and *will* to [5.4.1.1.VIII], it becomes obvious that neither the intrinsic nor the extrinsic senses of *will* fit into the context of the item. Rather, the immediate context very clearly shows that the situation described calls for *must* in its extrinsic sense. Within the sentence the

item is part of, a judgement on the side of the producer of the utterance expressed explicitly with the help of the verb *to believe*, which implies a certain opinion about the correctness or probability of the proposition expressed in the *that*-clause. Accordingly, if test takers are familiar with the fact that *must* can express exactly this (extrinsic) sense, whereas *will* cannot, the identification of alternative three as incorrect could be a fairly easy-to-achieve first step in finding the correct solution to the item.

As the analysis of the error rates for item [5.4.1.1.VIII] showed that the identification of *will* as an incorrect alternative did not seem to be problematic for most of the test takers, only one out of 34 test persons, or 3% of the testing population, decided for this distractor. The second distractor, alternative one, however, seemed to be relatively attractive for a larger group of test takers, i.e. 21% of the total test population for text four decided for this alternative. Comparing the correct solution to the item with alternative one, a focus on test taker's competence in the use of modal verbs expressing different time references becomes notable. Besides providing evidence for the use of extrinsic *must* in [5.4.1.1.VIII], the immediate context of the item as well as the context provided by the whole text also clearly implies that the fact, the likelihood of which is being judged with the help of extrinsic *must*, lies in the past. This can most clearly be seen if one takes a look at the tenses used within the clauses preceding the item. Apart from the main clause which introduces the believe expressed in the sentence, the two relative clauses contained in it use simple past verb tenses. As it is one of the most defining characteristics of the past tense in English that a respective "event/state" referred to with the help of a verb in the past tense "must have taken place in the past, with a gap between its completion and the present moment" (Quirk et al. 1985: 183), the situation described by the immediate context of the item can clearly be identified to lie in the past. As it is this situation which is judged according to its probability with the help of extrinsic *must*, the verb needs to express a past reference as well. Interestingly, the past of extrinsic modals is usually expressed with the help of *have* followed by the past participle of the main verb (Palmer 1990: 64). Although such constructions are very close in their formal structure to present perfect tense forms, it needs to be maintained that they can only be used to express references such as those typically expressed by the simple past. They do not normally include time references implied in the use of the present perfect of full verbs as such. [5.4.1.1.VIII] can be described as focusing on the particularities of tense usage with modal verbs. The potential of modals to have extrinsic as well as intrinsic meanings which is also focused on by the item is furthermore related to the way different time references are expressed by modal verbs. Whereas intrinsic *must* "has no past form" (Palmer 1990: 79), past

tense constructions with *have* and a past participle of the main verb "are freely used, however, with extrinsic modal meanings" (Quirk et al. 1985: 235), such as those of *must*. [5.4.1.1.VIII], thus not only aims at measuring test takers' competence in the use of the past tense with modal verbs in general, but also those aspects of grammatical competence which are related to its use with different senses of modals. The need for must in its extrinsic sense can relatively explicitly be inferred from the context of the sentence [4.1.1.1.VIII] is part of, particularly the verb form *to believe*. The same can be said about the need for a past form of the modal verb in question. Nevertheless, as was the case for [5.4.1.1.VI] and many other examples in the WBPT, the context provided by the whole text has an important supporting function. In the case of [5.4.1.1.VIII] the whole of text four frequently refers to past events and situations which might serve as an important cue for recognizing the need for a past form as the most appropriate solution to the item. Aspects, such as the ones just discussed that are related to the impact of contextual information on finding correct or appropriate solutions to multiple choice test items will also form an integral part of the following section which will be dealing with the way prepositions are tested in section one of the WBPT.

5.4.1.2 Testing Prepositions

In sum, all the four texts in section one of the WBPT contain 16 items focusing on the use of prepositions in English. One of the weak points of section one is the fact that the number of items focusing on prepositions is rather unevenly distributed across the four texts. Whereas text one only contains one item dealing with prepositions, texts three and four contain four such items, respectively. For text two, the figure is seven. It might therefore be argued that the four texts in section one cannot be considered as equivalent measures of the same grammatical contents. Whereas, the error ratios for tense items in section one were relatively diverse, ranging from items with an error ratio of under 20% to items which were completed incorrectly by over 80% of the testing population, the preposition items in section one yielded more consistent error mistake pattern. For seven out of 16 of these items the error ratio was between 53 % and 62%.[48]

One of the items with a fairly high error ratio is item 5 of version two:

[5.4.1.2.I]

Brenda (4) <u>was named</u> after US country singer Brenda Lee and began to perform _____

---

[48]Three items yielded lower error ratios, between 22% and 35% and five of the items were completed incorrectly by only a very small number of test persons, therefore, the percentages are not given here.

1) at
2) with
3) --five years old.

The main purpose of the item is to check test takers' understanding of the use of the preposition *at* with temporal meaning. According to Quirk et al. *at* can, amongst other things, be used as a preposition "of 'time position'", i.e. it can be used to refer to "points of time" (1985: 687f.). As is the case for many prepositions which have temporal meanings, the time reference of *at* can be assumed to be directly related to the dimensional sense of the preposition. In this sense, *at* normally occurs in PPs denoting a particular "simple position (or static location)" (675). As a preposition of position, *at* refers to a particular point in space treating the NP that it accompanies "as a dimensionless location", i.e. it does not bear implications of line or surface dimensions or even dimensions of an area or a volume (674). The same implications of meaning can be found in temporal uses of *at* where the word frequently refers to more or less fixed time points, such as "clock-time" as opposed to the duration of events or their extension into the future etc. (c.f. 687-692).

The situation provided by [5.4.1.2.I] puts a special emphasis on the point of time when Brenda Fassie started performing. Therefore, the item clearly implies that *at* in its temporal sense would be an appropriate alternative, especially if one considers the verb form of the clause containing the item. The verbal expression *to begin to do something* refers to a particular point in time, namely that when a certain action starts. Therefore, the selection of alternative one as the most appropriate solution to the item should not be problematic for a test taker who is able to recognize the temporal relation which is provided by the immediate context of the item and familiar with the common uses of *at* in its temporal sense. A closer look at the error rates for [5.4.1.2.I], however, reveals that the item was relatively problematic for more than half of the testing population: 55% of the total number of test takers answered the item incorrectly. Interestingly, the choice of incorrect alternatives is distributed evenly between distractors two and three, i.e. 50% of all test takers who answered the item incorrectly or 27% of the total testing population chose these alternatives, respectively.

One of the reasons for the high number of test persons who were not able to identify *at* as the correct solution to [5.4.1.2.I] might be the fact that the use of the preposition with reference to the age of a person does not seem to be the most common of its possible utilisations. The most common temporal meaning is that "introducing the time at which an event happens" (*OED*: s.v.). In the case of [5.4.1.2.I], the use of *at* as

referring to "the age at which one is" (ibid.) can be assumed to express the sentence subject's age with reference to a particular point in time for which this age has a certain relevance.

Relations between a person's age and a particular point of time or an action which happens or starts at this point of time are most commonly expressed in German with the help of the preposition *mit*. Accordingly, the fact that half of the test takers who answered the item incorrectly chose *with* as the most suitable answer could to a certain extent be explained by the fact that the majority of the test takers have German as their mother tongue. The preposition *with* in English has a number of senses including "pervasive" (Quirk et al. 1985: 684) and "instrumental meaning" (699) as well as implications of "solidarity or movement in sympathy"(702) or references to certain types of possession, i.e. the "notion of 'having'" (704). German "*mit* most often equates to English 'with'" (Joyce, n.d.) and can be said to have all of the possible meanings of *with* cited above. The fact that *with* can very often easily be translated by German *mit* and the fact of their high graphological as well as phonological similarity might be some of the reasons why so many test takers selected alternative two as solution for the item.

Alternative three, the use of the expression *years old* in combination with a number is very frequently used in English to refer to "a person's age" (*OED*: s.v.). However, it can be assumed that the situation provided by the context of the item puts a special emphasis on a certain point in time, the age of the subject is important for. Therefore, the main focus is on one point within the time period of Brenda Fassie's being *five years old* and not on her age in general as would be implied in the expression being used without the preposition *at*. An investigation of occurrences of *years old* in combination with a number and without a preposition within the British National Corpus (BNC) revealed that the expression is most frequently used in combination with the verb *to be* as can be seen by the following examples:

> [5.4.1.2.II]
> *He was 68 years old.*
> (BNC: A05 1419)

> [5.4.1.2.III]
> *He is a schoolteacher, 29 <u>years old</u> -- the age of Christ at Calvary, whose name is often in his mouth, averse though he is to "deities", and perhaps of Hamlet, whose words enter the novel.*
> (BNC: A05 1585)

The verb *to be* implies that the expression is used with reference to a general quality of, in the case of [5.4.1.2.II] and [5.4.1.2.III], persons as opposed to certain points in time when this quality is in the centre of attention.

The same can be said to be true of the verb *to feel* which is one of the few verbs, apart from *to be*, which occurs in combination with the expression within the BNC:

[5.4.1.2.IV]

And I also felt about twelve years old.

(BNC: AB3 387)

Like *to be*, the verb implies a certain general state which is not related to a certain point in time. If the expression *years old* in combination with a number is related to a certain point or period of time, this is frequently done with the help of the conjunction *when* introducing the sub-clause containing the expression.

[5.4.1.2.V]

That was in 1954 when he was 30 years old.

(BNC: A5J 25)

It needs to be maintained, however, that reference to a specific point in time is often most clearly expressed with the help of *at* in combination with *years old* and a number (c.f. *OED*: s.v.). With presenting alternative three as a distractor, item [5.4.1.2.I] can therefore be described as focusing on test takers' recognising of the importance of a particular point of time for the situation provided by the immediate context. This is rather explicitly expressed with the help of the verb *to begin*. Furthermore, by presenting alternative two as the second distractor, it focuses to some extent on problem areas with reference to the use of prepositions that can be assumed to be a result of German-English interference problems. From this it can be seen that the immediate context of the item, i.e. the sentence it is part of seems to contain the most important pieces of information necessary for finding the appropriate solution for [5.4.1.2.I].

Item 24 of version two focuses on the use of prepositions, or, more precisely, prepositional adverbs in combination with a verb:

[5.4.1.2.VI]

On Sunday evening, she died, leaving _____

1) back

2) <u>behind</u>

3) out

a son, Bongani.

The item was answered incorrectly by 45% of all test takers who were provided with text two in section one of the WBPT. The great majority, 93% of those subjects who chose one of the distractors, or 42% of the total testing population, decided for alternative one. Alternative three was selected by only one out of 33 test takers which accounts for 3% of the total number of subjects.

This item can be described as primarily focusing on test persons' knowledge of phrasal verbs. Phrasal verbs are described by Quirk et al. as "combinations of verb + adverb" which "behave syntactically or semantically as a single unit (1985: 714). Phrasal verbs are considered to be a type of "multi-word verb" (1161). They need to be distinguished from prepositional verbs as another member of this group. "A prepositional verb consists of a lexical verb followed by a preposition with which it is semantically and syntactically associated" (1155).[49] Although [5.4.1.2.VI] is discussed under the heading 'prepositions', it needs to be underscored that all the alternatives given for the item are not prepositions in the strict sense of the word, but prepositional adverbs. The category of prepositional adverbs is best described as "a particle which is formally identical to or related to a preposition" (713). In contrast to prepositions, prepositional adverbs are "capable of standing alone as an adjunct, conjunct, postmodifier etc. without the addition of a prepositional complement", i.e. a prepositional adverb "shares the form, but not the syntactic status of a preposition" (ibid.). One characteristic that prepositions and prepositional adverbs share, however, is the fact that both "commonly appear in idiomatic combination with the preceding verb" (714). Both groups of words are treated in the following paragraphs under the headline 'prepositions' because they share a number of features. Nevertheless, it needs to be

---

[49]For a detailed account of the contrast between phrasal and prepositional verbs, c.f. Quirk et al. (1985: 1152-1160) and Palmer (1988: 217-222).

clearly kept in mind that the two categories also show a number of, especially syntactic, differences and that they are in no way to be treated synonymously.

The choice of the correct solution to [5.4.1.2.VI] as well as alternative three result in the formation of the phrasal verbs, *to leave behind* and *to leave out*, respectively. These are used transitively with the NP *a son, Bongani* as their direct object. The choice of both alternatives results in the formation of a grammatically correct clause. What is important, however, for test takers in order to identify alternative three as incorrect or unsuitable in the context provided by text two is the fact that there exist considerable differences in meaning between the phrasal verbs to *leave behind* and *to leave out*: According to the *Longman Dictionary of Phrasal Verbs (LDPV)*, *leave behind* has the meaning of to "go away without" or "to go away with (something or someone) remaining behind (one)" (s.v.). In contrast, *leave out* is defined by the *LDPV* as the failure "to include" someone or something (s.v.). Accordingly, presenting two prepositional adverbs the choice of which results in the formation of two grammatically correct phrasal verb constructions as alternatives for the item can be said to put a special emphasis on test takers' understanding of semantic features of phrasal verbs. It is often claimed that knowledge of phrasal verbs is an aspect of lexical or vocabulary rather than grammatical competence (c.f. Palmer 1988: 216). [5.4.1.2.VI] can be described as measuring "knowledge of the meaning of certain words as well as the patterns and collocations in which they occur" which is a defining characteristic of vocabulary tests (Heaton 1988: 9). Nevertheless, it can be noted that grammar tests frequently include items which deal with phrasal verbs and their semantic aspects. One reason for this inclusion of vocabulary knowledge can be said to result from the fact that present day approaches to language proficiency, such as Bachman's CLA, clearly view vocabulary knowledge as one component of grammatical competence (1990: 87).[50]

The semantic opposition between the phrasal verbs *to leave out* and to *leave behind* did not, however, seem to be highly problematic for the vast majority of test takers. Rather, the error rates for the item imply that more than 40% of the testing population were not able to recognize differences between the prepositional adverb *back* and the correct solution *behind*. Alternative one is the only alternative to the item which does not lead to the formation of a phrasal verb or an acceptable verb collocation in general. Again, it needs to be maintained that alternative one primarily concentrates on semantic as opposed to grammatical, or more precisely syntactic, aspects related to the use of prepositional adverbs. According to the *OED*, *back* in its adverbial sense,

---

[50]A certain amount of items which focus on vocabulary knowledge can for instance also be found in the grammar sections of the TUC-PT.

refers to "a direction to the rear", or literally, to "the direction of one's back, or the back of any object in question" (s.v.) Most of the other directional meanings of *back* also imply this particular backwardness (s.v.) which is not a primary part of the semantics of the other alternatives provided for the item. In its primary directional sense *out* as an adverb expresses "motion or direction from within a space, or from a point considered as a centre" (s.v.). It can therefore be assumed that the word frequently refers to the idea of a motion originating from inside a particular space or object, a semantic aspect that can also be traced in the phrasal verb to *leave out* discussed above. This aspect is not included in the semantics of alternative two, *behind*, which commonly refers to the position "in a place whence those to whom the reference is made have departed" (s.v.). The *OED* furthermore makes a reference to the verb *to leave* often occurring in combination with *behind* as well as to the specific figurative sense of *behind* as referring to a person's death: "In the position, condition, or state which a person or thing has left, e.g. in existence after a one's death" (ibid.).

This particular use of the adverb in combination with a verb hints to the high significance of idiomaticity that is an aspect of the semantics of most phrasal verbs. It can be seen that a combination of the prepositions *behind* as well as *out*, but not *back*, with the verb *to leave* leads to meanings that differ to various extents from the meanings of the individual parts of the construction. It can therefore be assumed that [5.4.1.2.VI] focuses for instance on the understanding of and the competence in the semantics of the idiomatic meanings of *to leave out* and *to leave behind*. With reference to the concept of idiomatic meanings of phrasal verbs, Palmer points out that

> the issue of idiomaticity is closely tied up with 'transparency' or literalness [...]. The meaning of a combination [of verb and preposition or prepositional adverb or both; K.U.] can be said to be transparent (or literal) if it can be deduced form the meaning of the individual parts [...]. If it cannot, it is opaque (1988: 217).

It needs to be emphasised, however, that the division between transparency and opaqueness is not in all cases clear-cut. Rather, there exists "no clear dividing line", or a continuum "between idiomatic and non-idiomatic" uses of phrasal verbs (ibid.). For [5.4.1.2.VI], it can be argued that the choice of the correct alternative *behind* as well as the distractor *out* leads to a construction of phrasal verbs the meaning of which lies on some point within this continuum. It can be assumed that the meaning of both expressions, though idiomatic to some extent, could be identified if test takers possess a suitable amount of knowledge of the meanings and uses of the prepositional adverbs *behind* and *out*. Choosing *out* as a distractor did not prove to have a highly discriminating effect as the alternative did not seem to be attractive to the broad majority of subjects who answered the item incorrectly. It can therefore be assumed that

most testees possessed sufficient knowledge of the meaning of either the phrasal verb *to leave out* or were able to reconstruct the meaning as not suitable on the basis of the meanings of the components of the phrasal verb. In contrast, the second distractor *back* was chosen by a relatively high number of test persons. It can, however, be clearly identified as incorrect in the context in question as the notion of somebody being in a certain state or position after someone else has left which is one possible aspect of the meaning of *behind* (c.f. *OED*: s.v.) is not included in the semantics of the adverb *back*. It can therefore be assumed that [5.4.1.2.VI] does not only focus on knowledge of phrasal verbs as purely idiomatic constructions but also on competence in the use of prepositional adverbs as such. With reference to the influence of contextual features on the item it can be said that the context of the sentence surrounding [5.4.1.2.VI], especially the verb form *to die* functions as a major cue for the identification of *behind* as the most appropriate solution as this verb form clearly hints to the metaphorical or figurative of *to leave behind* mentioned above. Nevertheless, the context of the whole of text two, which has the death of Fassie as one of its central topics can be described to add further pieces of valuable information.

Like [5.4.1.2.VI], item nine of text version focuses on the use of prepositional adverbs within phrasal verb constructions. Whereas for [5.4.1.2.VI] it could under certain circumstances be assumed to be possible for test takers to identify the correct solution on the basis of sufficient competence in the uses and meanings of prepositional adverbs, aspects of idiomaticity seem to play a very crucial role for the item to be discussed in the following section:

[5.4.1.2.VII]

These rocks have been an important part of formal rituals carried

1) on

2) <u>out</u>

3) through

by the people who came there.

The error rates for the item are considerably high: 68% of a total testing population of 34 test takers decided for an unsuitable alternative. 59% chose alternative one; 9% decided for distractor three, i.e. the figures are rather unevenly distributed. As was the case for [5.4.1.2.VI], [5.4.1.2.VII] mainly focuses on combinations of verbs with prepositional adverbs, not prepositions as such.

The selection of the prepositional adverbs *on* as well as *out* results in the formation of a phrasal verb, the selection on the preposition *through* does not. Alternative three was chosen by a relatively small number of testees, its discriminating value can therefore be judged relatively low. The alternative can be said to basically focus on German-English interference problems. The context of the text which deals with unusual rock formations, which seem to have been the basis for certain rituals, and particularly the context of the sentence surrounding the item imply the need for a verbal construction which is close in meaning to German *durchführen*. Especially the NP *formal rituals* which is more closely defined by the participle clause the item is part of collocates relatively well with a verb form expressing this meaning. The fact that English *through* can in many cases function as a translation equivalent for German *durch* might have led 14% of the target group to chose this distractor. The basic reason for providing *through* as an alternative to the item was to account for differences between German and English with reference to certain collocations or idiomatic expressions which cannot be translated word by word.

This difference did however not seem to be highly problematic for the majority of subjects. Rather, most test takers seemed to have considerable difficulties with idiomatic verb collocations within the English language as such. The fact that 56% of the total target group decided for alternative one which resulted in the formation of the phrasal verb *to carry on* used here in a participle construction implies that the majority of test takers was not sufficiently familiar with semantic differences between this construction and the correct solution *to carry out* (*carried out*). It needs to be maintained that the item primarily focuses on aspects of idiomaticity which are clearly a component of vocabulary competence. Both the phrasal verbs *to carry on* and *to carry out* are described by the *LDPV* as idiomatic as opposed to literal expressions (s.v.). As has been mentioned earlier on, the concept of idiomaticity can be described as a type of continuum with clear-cut divisions between absolute literalness and absolute idiomaticity being difficult to draw (c.f. Palmer 1988: 217). The degree of idiomaticity for *to carry on* and *to carry out* can however be judged to be relatively high when compared for instance to the phrasal verbs *to leave behind* and *to leave out* discussed above. Whereas for the latter two test takers could under certain circumstances be assumed to be able to detect their meanings on the basis of sufficient knowledge of the prepositional adverbs *behind* and *out* as well as the verb *to leave*, this does not seem to be the case for the *to carry on* and *to carry out*.

The basic emphasis of [5.4.1.2.VII] therefore lies on the assessment of test takers' familiarity with the meanings of particular phrasal verbs. According to the

LDPV, *to carry on* can have the literal sense of taking "(something) on", e.g. a plane, as well as an idiomatic sense expressing the continuation of a certain activity (s.v.). The correct alternative to the item can also have literal as well as idiomatic senses. The literal meaning of *to carry out* is "to lift (something or someone) and take it out" (s.v.). Clearly [5.4.1.2.VII], focuses on the idiomatic sense of the word which is in fact a translation equivalent to German *durchführen*. In this sense *to carry out* means "to fulfil or perform (something)" (ibid.). Comparing the idiomatic and the literal sense of the expressions it becomes obvious that the meaning of the former can hardly be identified on the basis of the meanings of the components of the respective phrasal verb. It needs to be maintained that items such as [5.4.1.2.VII] hardly allow any general abstraction about a test taker's competence in the meaning and use of prepositions and/or prepositional adverbs as it primarily focuses on competence in the semantics of a very limited set of phrasal verbs. The focus of the item is therefore a rather restricted one. With reference to the importance of the context for being able to solve the item it becomes notable that the immediate context of the sentence, especially the NP *formal rituals* influences the choice of a particular alternative rather than the context of the whole text. It could therefore be supposed that the item would have yielded the same or nearly the same error rates when being presented with as little context as one single sentence. This practise would however imply a number of disadvantages which will be further discussed in chapters 5.4.2 and 5.4.2.3 below.

Whereas for [5.4.1.2.VII] it is first of all semantic or lexical knowledge that is focused on, item ten of version three puts an emphasis on semantic as well as syntactic aspects.

[5.4.1.2.VIII]

But _____

1) out of

2) outside

3) outwards

the west and the Great Plains, far fewer schools will benefit (11) from these changes.

The item focuses on competence in the use of prepositions as opposed to adverbs. It was answered incorrectly by 59% of the total testing population. 31% chose alternative one and 28% decided for alternative three. All three alternatives to the item include semantic aspects of a certain remoteness, two of them, one and two, are prepositions, alternative three is an adverb. Therefore, one of the main difficulties of the item is to

decide for the correct word class that fits into the context provided. Furthermore, test takers need to realize the, sometimes relatively slight, meaning differences that exist between the correct alternative and the distractors.

Although all three alternatives share the semantic component of remoteness, their meanings nevertheless seem to treat this remoteness in varying ways. The preposition *outside* generally denotes a particular position, carrying meanings such as "on the outer side of; external to" *(OED*: s.v.), in particular, it is often used to refer to positions that are located "beyond the limits of (any domain of action or thought, any subject or matter)" (ibid.). A second possible sense of the preposition *outside* is that of "motion or direction: To the outer side of, to the exterior of, to what lies without or beyond" (ibid.). The second preposition among the alternatives for [5.4.1.2.VIII] has meanings that are closely related to those of *outside*. However, *out of* seems to be used primarily with reference to "motion or direction" *(OED*: s.v.) which is not the case for *outside*. In the sense just mentioned, *out of* refers to a "motion or direction" that originates "from within (a containing space or thing)" (ibid.). The word may, nevertheless also have a positional meaning, referring to a location which is "not within (a space or containing thing)" or which is "beyond the confines of" (ibid.). The preposition *out of* "may express the position resulting from the motion" included in the directional sense of the word, "or that" resulting from the opposite of "inward motion" (ibid.). From this it becomes obvious that the semantics of the preposition *out of* mostly include a strong direct or indirect reference to a motion of whatever kind, although it can, under certain circumstances, also have positional meaning.

As both, alternative one and two are prepositions and as such have the defining characteristic to precede or complement NPs, the decision between *out of* and *outside* needs to be made more or less exclusively on the basis of competence of the possible meanings of prepositions. As mentioned in the task specification, test takers are not expected to decide for the one and only correct alternative but rather for the most appropriate one. Taking into account, the relatively high significance of aspects of motion for the meaning of *out of* which does not seem to be included to such a high extent in the semantics of *outside,* it is justifiable to describe alternative two as more appropriate within the context of the item than alternative one. Nevertheless, it is not possible to describe alternative one as entirely incorrect. The fact that *out of* was chosen by more than half of the subjects who did not decide for the alternative which is correct according to the answer key, calls for an incorporation of the alternative into the key. Although there exist slight differences with respect to the different semantic emphases included in the prepositions *outside* and *out of*, the analysis of the error rates showed

that these are probably too slight to be recognized by the target group. Therefore, it needs to be maintained that although *outside* would be the more appropriate solution to the item, the selection of *out of* would also be acceptable in the context of [5.4.1.2.VIII].

Whereas the decision between alternative one and two seems to be one of varying degrees of appropriacy of meaning, alternative three clearly focuses on test takers' understanding of the syntactic properties of adverbs as opposed to prepositions. The immediate context of the item consists of an NP denoting particular places, i.e. the American west and the Great Plains. The clause that follows the NP in question refers to a certain situation that is characteristic for the places mentioned in the NP. Therefore, in order to be able to identify alternative three as incorrect, test takers need to recognize that the construction calls for a preposition, specifying a certain position with reference to the places mentioned in the NPs *the west* and *the Great Plains*. The fact that the item calls for the selection of a preposition can most clearly be understood if test takers are aware of the most distinctive characteristic of prepositions to take NPs as complements.

According to Quirk et al. "the prepositional complement is characteristically a noun phrase, a nominal *wh*-clause, or a nominal *-ing* clause" (1985: 657). In contrast to alternatives one and two, alternative three, the adverb *outwards* cannot take an NP or the clause types mentioned by Quirk et al. as a complement. Huddleston and Pullum define adverbs as "a grammatically distinct category of words whose members are characteristically used to modify verbs but not nouns" (2002: 562). Although this is a quite general definition of the term which does not take into account the different possible functions of adverbs (c.f. Quirk et al. 1985: 439f.) as well as the formal and semantic similarities that exist between adverbs and prepositions (662) it expresses a very important syntactic characteristic that distinguishes adverbs from prepositions: Whereas prepositions characteristically accompany nouns or noun phrases, adverbs do not. Having recognized the fact that alternative three is an adverb and can therefore not be used in combination with the NP following the item, alternative three should be identifiable as unsuitable in this context. This identification of *outwards* as incorrect is, in contrast to decisions called for by items such as [5.4.1.2.VI] or [5.4.1.2.VII] more strongly based on the syntactic properties of the alternative. Nevertheless, it needs to be kept in mind that the identification of alternative three as incorrect could also be achieved on the basis of a detailed knowledge of the possible meanings of the adverb *outwards*. As mentioned earlier on, the NP forming the immediate context for [5.4.1.2.VIII] denotes a particular position. *Outwards* can, however, only denote "direction but not position" (Quirk et al. 1985: 516).

Therefore, it can be assumed that although the item puts a greater focus on syntactic competencies regarding the use of prepositions and adverbs than a broad number of the other preposition items in section one of the WBPT, vocabulary/lexical competence also plays a highly significant role for finding the most appropriate/an appropriate alternative to the item. As can be seen from what has been mentioned above, the immediate context of the sentence surrounding [5.4.1.2.VIII], its semantic implications as well as syntactic structure play a very important role for the identification of the (most) appropriate alternative to [5.4.1.2.VIII].

Whereas error rates were relatively evenly distributed across the distractors for [5.4.1.2.VIII], for item 25 of version three, a different picture emerges.

[5.4.1.2.IX]

"You've got the same kids _____

1) <u>for</u>

2) since

3) up to

five years, so you can't repeat anything – you've got to come up with new stuff."

[5.4.1.2.IX] was answered incorrectly by 22% of a total testing population of 32 subjects which is a relatively low percentage when compared to error rates for the majority of the other items focusing on prepositions in section one of the WBPT. Nevertheless, the fact that more than one fifth of the test persons had problems with the item implies that problems with the prepositions focused on by [5.4.1.2.IX] cannot be neglected. Not surprisingly, the basic problem arising out of the item seems to consist in the use of the preposition *for* as opposed to *since*. The whole 22% of the target group which chose a distractor decided for alternative two, i.e. *since*. The third alternative, *up to*, was not chosen by any of the subjects which makes the usefulness of *up to* as a distractor in the context of [5.4.1.2.IX] questionable.

The fact that more than one fifth of the testing population had problems identifying *for* as the correct alternative might to some extent be explainable by the variety of meanings that *for* as a preposition can possess. As pointed out by Huddleston

and Pullum, "*for* is perhaps the most polysemous of the prepositions of English, with a plethora of subtly distinct meanings and a small set of grammaticised uses in addition" (2002: 655). The word can, for instance, refer to a certain cause, a reason, or a purpose (Quirk et al. 1985: 695). Some researchers describe this meaning of *for* as one of its "core functions" (Lindstromberg 1997: 223). Another, primary meaning of the word, which is focused on by [5.4.1.2.IX], is that of temporal duration, which needs to be clearly distinguished from the concept of "time position" (Quirk et al. 1985: 689). In contrast to the preposition *since*, which generally refers to a point in time, *for* used in a temporal sense denotes a "stretch of time" (691). This is also reflected in the *OED* which describes the word's temporal meaning as one "of duration and extension" (s.v.). As can be seen from [5.4.1.2.IX], the preposition *for* in this sense, often takes "phrases of time measurement" as its complement (Leech & Svartvik 1994: 87).

The potential to take "a temporal noun phrase" as a complement is, however, not only a common feature of the preposition *for* but also of *since*[51]. The syntactic properties of the words can therefore be assumed to be more or less similar. Semantically speaking, the reference to time expressed in the NPs commonly accompanying *for* and *since* does, however, differ to a considerable extent and it is this particular difference which is primarily focused on by [5.4.1.2.IX]. In contrast to *for*, *since* is "almost exclusively" used with reference to time (Quirk et al. 1985: 691). Like that of *for*, its meaning includes references to a particular duration of time, which can be seen from the following definitions of the word provided by the *OED*: "Ever or progressively from (a specified time, etc.) till now" and "during the period between (a specified time) and now" (s.v.). These definitions, however, also imply that the word expresses a main emphasis not on the actual duration of the time period in question but on the particular point in time when this time period started. Therefore, one primary aim of [5.4.1.2.IX] consists in the assessment of test takers' ability to recognize semantic differences between the prepositions *for* and *since*. The fact that none of the test persons chose *up to* as a solution to the item, implies that, although it was intended to, the item does in fact not focus on meaning differences between this preposition and the other two alternatives. Therefore, the item can be said not to include a characteristic that multiple choice test items should ideally possess. As claimed by Alderson et al. each incorrect alternative to such items should have a certain attractiveness for at least smaller groups of test takers (1995: 48). All of the test persons seem to have easily recognized that, although the word shares the feature of *for* and *since* to refer to a certain time duration, the perspective from which it is expressed by *up to* is different

---

[51]If not otherwise indicated, the remainder of this chapter will refer to *for* and *since* as prepositions only.

from that of *for* as well as that of *since*. In general, *up to* does not refer to the actual duration of a period or its starting point but rather to its end (c.f. *OED*: s.v.).

The main difficulty of the item, therefore, seems to consist in the distinction between the use of *for* and *since*. Apart from measuring test takers' competence in the semantics of *for* and *since*, [5.4.1.2.IX] can also be said to focus on their understanding of the use of the words in connection with verb tenses. Whereas *for* can be used with a number of different tenses, such as the simple present or the simple past, *since* is generally used with the present perfect and a particular time expression (c.f. Quirk et al. 1985: 194). One of the main uses of the present perfect tense is that which relates "a past event/state to a present time orientation" (192). Expressing a certain relation to the present or the TU is also a primary characteristic of the temporal meaning of *since*, which can clearly be seen from the *OED*'s definitions of the word cited above. Though possible, such an orientation can not be said to be as typical for *for* as it is for *since*. From the verb form of the clause [5.4.1.2.IX] is part of, it must be fairly obvious for a test person who is familiar with the relations of *for* and *since* to the use of tenses that *since* is an unsuitable alternative for [5.4.1.2.IX]. The clause surrounding [5.4.1.2.IX] as well as the clauses following it employ simple present verb tenses which makes *since* a rather unsuitable selection as there is no past reference expressed within the sentence. As far as the importance of the context of the whole of text three for finding the most appropriate alternative to [5.4.1.2.IX] is concerned it needs to be maintained that for this item the contextual information provided on the level of the sentence seems to be sufficient to definitely determine alternative one as the most appropriate or correct solution.

Apart from the correct solution being relatively unambiguously identifiable even within the restricted context of the sentence, the distinction between *for* and *since* remained problematic for a fairly high number of subjects. Besides the factors just mentioned, differences between German and English seem to be a reason for the problems arising out of [5.4.1.2.IX]. Apart from the semantic and syntactic properties of prepositions within the English language, the item also focuses to a certain extent on test takers' ability to distinguish typical uses of prepositions within their mother tongue from their use in the target language. In particular, some uses of the German preposition *seit* are often incorrectly translated by learners with English *since*. German *seit* can be used with reference to the starting point of a time period as well as to the duration of this period in general. Therefore, the different meanings of *for* and *since* in English can, under certain circumstances, be both expressed by *seit*. The word can generally be defined to express "daß ein Geschehen in einer Zeitspanne verläuft, die in der

Vergangenheit begonnen hat und bis zur Sprechergegenwart reicht" (Schröder 1986: 166f.). However, even if L1 interferences may influence the majority of the test persons in their choice for a certain alternative, it needs to be kept in mind that even in German the context provided by the sentence surrounding [5.4.1.2.IX] would make the use of *seit* impossible as it does not make any reference to a past event or situation. Therefore, [5.4.1.2.IX] does not directly focus on problems arising out of differences in the grammatical structures of test takers' L1 and their L2 although it focuses on the use and meaning of prepositions which seem to be especially problematic for German learners of English. As has been mentioned earlier on, the context of the sentence, particularly the tense forms used in it, play a crucial role for determining alternative one as the correct solution to the item.

5.4.2 Multiple Choice Gap Filling on Text versus Sentence Basis: Comparison of Findings with Sections 2.1 and 2.2 of the TUC-PT

As has been mentioned before, contextual features play a fundamental role for multiple choice test tasks. Apart from providing the testee with necessary linguistic and factual background information, it can for instance also have a crucial impact on the particular language contents that are being tested. With reference to the testing of grammar, Heaton points out that "the provision of a detailed context", e.g. in the form of "a paragraph or series of paragraphs of descriptive, narrative or expository prose" can restrict "the range of grammatical features being tested" (1988: 36). Tests of grammar which attempt to test as wide a variety of grammatical contents as possible therefore sometimes use more or less isolated sentences, which contain the respective phenomenon, as input material. It needs to be maintained that some aspects of English grammar are in fact very difficult to test on the basis of a longer, coherent text. As an example for the possible restrictions that different types of texts can impose on the grammatical contents to be tested, Heaton mentions that "it is usually impossible, for example, to test the future continuous tense, in a narrative set in the past (unless direct speech is used)" (ibid.). Apart from all these possible restrictions that the provision of an extended context for multiple choice test tasks may imply, the advantages of presenting such items in context seem to outweigh the disadvantages. Using coherent and possibly also authentic text material first of all has the advantage of reflecting actual real-life language use to a much higher extent than isolated sentences, as natural communication normally consists of successions of utterances which bear a variety of interrelations with each other. "Furthermore", as has been outlined in chapter 5.3 above, the provision of context helps to decrease test ambiguities. In contrast, "short

decontextualised sentences can lead to ambiguity as they are usually open to several interpretations when used as stems for multiple-choice items" (ibid.).

The grammar component of the Chemnitz University of Technology English Language Placement Test (TUC-PT) contains two sections which employ a multiple choice testing format based on single sentences. These are section 2.1 and 2.2 of the test. Section 2.1 consists of ten gapped sentences or combinations of short sentences. The testees are required to choose the correct out of four alternatives that are given below each of the sentences. Section 2.2 contains 14 single sentences each of which contains three alternative words or expressions printed in italics. Here, the test taker's task is to underline the correct alternative in each sentence.

In the following, two examples for the testing of tenses from section 2.1 of the WBPT and two items out of sections 2.1 and 2.2 respectively which focus on the testing of prepositions will be analysed and discussed. Special importance will be given to the influence of the rather limited context on the items and the testing situation as such.

5.4.2.1 Testing Tenses

One example from the TUC-PT for the testing of tenses on the basis of a combination of two sentences, in this case an exclamation followed by a question, is item four of section 2.1.

[5.4.2.1.I]

Just look at the dog – he's completely covered in mud! What on earth <u>. . . he . .</u> ?

    a. was . . . doing   b. had . . . done

    c. did . . . do   <u>d. has . . . been doing</u>

Compared with all the other items in the TUC-PT focusing on English verb tenses, [5.4.2.1.I] yielded the highest error ratio, i.e. the item was answered incorrectly by 80 % of a total testing population of 137 test persons that have been analysed for the purposes of this paper. The selection of alternatives reveals that [5.4.2.1.I] focuses on the use of the present perfect progressive (alternative d) as opposed to the past progressive (alternative a), the past perfect (alternative b), and the simple past (alternative c). All of the three disctractors seemed reasonably attractive to some parts of the target group. Alternatives a and c were chosen by 17% and 14% of all test takers, respectively. The highest percentage of the target group, 49%, however, decided for alternative b, the simple past form of the verb *to do*.

Clearly, all four alternatives imply a certain temporal relation to the past. Furthermore, it is interesting to note that the main emphasis of the item seems to lie upon differences in the use of simple as opposed to progressive and/or perfective forms of the verb. As providing a small amount of context for multiple choice items is often described to increase the possibility for items to become ambiguous (c.f. Alderson et al. 1995: 47f.), the item will in the following be analysed with regard to the question whether the relatively restricted context provided is sufficient in order to determine alternative d, the present perfect progressive form of *to do*, as the only possible solution to be chosen. Quirk, Greenbaum, Leech, and Svartvik provide the following general definition of the perfect progressive. "The perfect progressive denotes a temporary situation leading up to the present moment. The progressive overtones of incompletion and emotional colouring can also be found" (1974: 97). Parts of this definition of the present perfect progressive are very frequently used in TEFL contexts in order to explain the use of the present perfect progressive which can for instance be seen if one takes a look at the following quotation from an English language textbooks.

> We use the [present perfect; K.U.] continuous to emphasize the length of time the activity has lasted, over hours, days, weeks, moths, or years leading up to the present.
>
> *The Rolling Stones **have been playing** together for over 30 years.*
> *I've **been doing** my homework for hours.*
> (Soars & Soars 1998: 17).

The expression 'up to the present' which is used by Quirk et al. (1974) as well as the textbook and especially the examples provided by Soars and Soars (1998) might create the impression that the present perfect continuous is only used to refer to situations which have a starting point in the past and which are not finished at the TU. From [5.4.2.1.I] however, it can clearly be inferred that the action to be denoted by the present perfect continuous verb form is definitely finished. This becomes obvious by the expression *Look at the dog – he's completely covered in mud!* The verb form to be supplied can thus not be described to include implications of incompletion. Rather, the exclamation preceding the question which contains a present perfect continuous form hints to a certain amount of excitement or "emotional colouring" (Quirk et al. 1974: 97) which can be described as one of the possible justifications for the progressive aspect to be used in this context. Much more importantly, however, the exclamation hints to a special interest in the actual process of the dog performing an action that left him covered in mud. This is most clearly expressed by the progressive aspect which "indicates temporariness", i.e. "an action in progress instead of the occurrence of an action or the existence of a state" (92). With a past reference, the latter would much

rather be expressed with the help of a simple past form, and it can be said that because this latter part of the definition provided above does not apply in the case of [5.4.2.1.I] alternative c, the simple past form of the verb, can relatively clearly be described as incorrect. The same is true for alternative b, the past perfect form of *to do*. This can very easily be explained by the fact that "past perfect verb phrases are context dependent", i.e. they refer "to a time before a past time signalled elsewhere" (Biber et al. 1999: 469). The small amount of context that is provided by the stem of [5.4.2.1.I] does in fact not contain any past time reference apart from that contained in the four alternatives given.

As has been mentioned above, the actual process of the dog performing a certain action, is very important for [5.4.2.1.I]. However, it can not be described as the only aspect to be expressed by the verb form needed to complete the item successfully. This is also the reason why it is possible to describe alternative a, the past progressive form of the verb, as an incorrect solution to [5.4.2.1.I]. The exclamation *Look at the dog – he's completely covered in mud!*, which has been cited above, plays a fundamental role for determining alternative d as the only possible solution to be chosen. It implies a very definite reference to the result of the action denoted by the verb form to be supplied. It furthermore shows that the immediate result of this action can be seen at the TU; this can for instance be explained by the fact that the verb tenses used in the exclamation are all in the present. Such a reference to the present time or the TU is most effectively expressed with the help of the perfective aspect. The most general characterization of the present perfect (or perfective) is the fact that it expresses "'past with current relevance'" (Quirk et al. 1985: 192), which can definitely be assumed to be true in the case of the situation expressed in [5.4.2.1.I]. This present reference of the action to be denoted by the verb form together with the focus on an action in progress mentioned above, rather clearly call for the selection of alternative d as the correct solution to [5.4.2.1.I].[52]

From the above discussion it can be seen that although [5.4.2.1.I] provides testees with a fairly limited amount of linguistic and factual context, all of the distractors could clearly be identified as incorrect which implies that Alderson et al.'s claim of a restricted context for multiple choice test items evoking potential ambiguity could not be proven for [5.4.2.1.I]. However, although the two sentences contain a number of features, such as *just* or *look* that might be associated with natural, oral communication, it can still be claimed that the format is to a certain extent problematic.

---

[52]Although the example discussed here contains semantic aspects of the progressive in combination with the perfective aspect it needs to be underlined that the perfective progressive cannot in all cases be described as a simple combination of features that are typical for its two components (c.f. Quirk et al. 1985: 211ff.).

This is the case because the impression might be created that language and/or communication can function in decontextualized bits which is essentially untrue. With reference to this problem, Heaton states the following. "Note that context is of the utmost importance in all tests. Decontextualized multiple-choice items can do considerable harm by conveying the impression that language can be learnt and used free of any context" (1988: 28). One of the reasons why [5.4.2.1.I] is presented with such a small amount of context could be the fact mentioned in chapter 4.4 that some grammatical constructions are very difficult to test within longer text material (c.f. Heaton 1988: 36). In a corpus-based analysis, Biber et al. found that "perfect progressive verb phrases are very rare" in English (1999: 462). The question can thus be raised whether the fact of a given grammatical construction being rare and difficult to find in broader language contexts justifies its being tested in isolation from a more natural context than that provided by single sentences. Apart from the problems just mentioned, it needs to be underlined that [5.4.2.1.I] can be described as a well structured and objective test item since there exists only one possible correct alternative which can clearly be identified on the basis of a detailed knowledge of grammatical rules related to the use of tenses in English. In other words, the item is constructed in such a way that even the fairly small amount of context provided does not lead to grammatical or semantic ambiguities.

On the basis of what has just been discussed it becomes possible to claim that, in order to be unambiguous, sentence-based multiple choice test items that focus on tenses need to be constructed in such a way that the context provided does not leave any space for alternative solutions apart from the correct solution(s) cited in the answer key. An example for an item which clearly shows that there seems to exists the need for this practise is item ten of section 2.1 of the TUC-PT.

[5.4.2.1.II]
"Tom told me yesterday that he _ . _ . to see us this morning, but it's already 2 o'clock and he still hasn't arrived."
a. comes   c. came
b. will come   d. would come

Although the content of the sentence contains a number of cues which hint to alternative d as the correct and only possible solution to [5.4.2.1.II], the item was answered incorrectly by almost one third of the target group, the precise figure is 28%. For most of these 28% of testees, alternative b seemed most attractive, it was chosen by 23% of all 137 test persons. A fairly small number of subjects, 4% and 1% decided for

87

alternatives a and c, respectively. Obviously, the item focuses on the use of tenses in reported speech constructions in general and on the principles of backshift in particular. Backshift in reported speech "can occur" if "the **tense** of the matrix clause is past" (Huddleston & Pullum 2002: 153). However, as has been mentioned in the context of the discussion of [5.4.1.1.VI] backshifting is a phenomenon that "is normally optional" (Quirk et al. 1985: 188). Yet, for backshifting in constructions with a reporting verb in the past to be optional, it is crucial "that the original utterance is still applicable and relevant" at the time it is reported (Huddleston & Pullum 2002: 156). As was shown with reference to [5.4.1.1.VI] for some reported speech constructions with a past reporting verb, it can be quite problematic to clearly determine backshifted verb forms in the reported clause as the only forms possible, even if the original text used one of these forms.

For [5.4.2.1.II] however, such problems are not notable. In total, the construction contains five linguistic clues, the combination of which make the use of backshift doubtlessly necessary. These are *told*, *yesterday*, *this morning*, *but it's already 2 o'clock*, and *he still hasn't arrived*. The reporting verb *told*, in combination with the time adverbial *yesterday*, illustrates the fact that the proposition to be reported was made at a point in time that is more or less remote from the moment of reporting. *This morning* implies a reference to a time later than the utterance made by Tom and further underscores this remoteness. In addition to that, the gap that exists between the two time points or periods referred to by *yesterday* and this *morning* shows that the proposition made by Tom included a reference to the future, namely the following morning. The clause *but it's already 2 o'clock* serves the function of further describing the time of reporting and of showing another, temporal, gap between the time the proposition or promise made by Tom was to be fulfilled and the actual situation that is described. In the clause *he still hasn't arrived*, the negative form of the VP explicitly shows that the proposition made by Tom the day before does not apply at the moment of its reporting. Relating all these properties of the construction to what Huddleston and Pullum define as a crucial condition for backshift to be optional, it becomes obvious that for [5.4.2.1.II] such optionality cannot be claimed. Through the use of a rather high number of contextual cues within one and the same sentence construction, it is doubtlessly emphasised that the necessary condition for backshift being obligatory is fulfilled. On the basis of this assumption, alternatives a and b can clearly be determined to be incorrect in the context of the item as both the simple present tense as well as a reference to future time with the help of *will* are not backshifted forms. Alternatives c and d, however, can both be examples of backshift of a simple present or a *will*-future

verb form, respectively. As mentioned before, a proposition that contained a future reference at the time it was first uttered, is reported in the context of [5.4.2.1.II]. "Future time can be expressed with the present tense or with the modal *will*" (Biber et al. 1999: 452). However, the use of the present with future reference most typically occurs "in subordinate clauses" or with regard to "scheduled events" (Greenbaum 1996: 257). This, however, cannot be assumed to apply to the proposition made by Tom which implies that the original utterance was most probably made using the modal *will*. Alternative d, the backshifted form of this way of establishing future reference can, thus, rather clearly be identified as the only possible solution to [5.4.2.1.II]. Like [5.4.2.1.I], the item is grammatically unambiguous which is basically due to the very high proportion of cues provided by the context of the sentence. Like [5.4.2.1.I], the item evokes the preliminary picture of natural, oral communication, it is, for instance, printed in question marks and contains a number of contracted forms. Nevertheless, the question remains, whether utterances containing such an unambiguous justification of the need for backshift are typical of natural communication in English. [5.4.2.1.II] can therefore be described as a further example for sentence-based multiple choice items which are constructed in such a way that they are highly objective but which have the obvious disadvantage of presenting a particular grammar point in isolation from a broader context that resembles authentic instances of language use.

In the following chapter, a closer look will be taken at the testing of prepositions in sections 2.1 and 2.2 of the TUC-PT. As was the case for the testing of tenses, special attention will be given to the testing of grammatical items within a relatively limited context and on questions concerning the impact of this limited context on test objectivity.

5.4.2.2 Testing Prepositions

Sections 2.1 and 2.2 of the TUC-PT contain a fairly low number of items focusing on the testing of prepositions. In total, the two sections include two such items which will be discussed below. The first of the items to be analysed is item six of section 2.1

[5.4.2.2.I]

Please finish this test . . . lunchtime.

a. until   c. by

b. to   d. till

89

In very general terms, [5.4.2.2.I] focuses on the use of temporal prepositions expressing a particular time limit. It was answered incorrectly by a very high proportion of the target group, namely 77%. Of the distractors, only alternatives a and d seemed attractive to the testees, alternative b was only chosen by one of the subjects (1% of the total target group). The error rates for alternatives a and d are 46% and 30%, respectively. It is interesting that all the distractors for [5.4.2.2.I] can often be used synonymously (c.f. *OED*: s.v. *to*). Furthermore, all four alternatives most commonly refer in one or another way to the end of a time period. However, the correct solution to [5.4.2.2.I], the preposition *by*, is the only one of the four alternatives "which co-occurs only with momentary verbs" (Quirk et al. 1985: 690), such as *to finish*. Thus, *by*, explicitly refers to "an end point" (ibid.), or, more precisely, "the time at which the result of an event is in existence" (692). In the case of [5.4.2.2.I], this time is expressed by the NP *lunchtime*, and the event mentioned by Quirk et al. can be described as the finishing of the test. It can be seen from the fact that the authors use the term event in the definition of *by* mentioned above that the preposition is only used with momentary verbs which generally refer to events as opposed to ongoing situations. From the semantics of the verb *to finish* it can be seen that it includes many aspects of momentariness as the event of finishing is generally momentary in nature. It is, thus, first of all the predicate of the sentence which explicitly defines alternative c as the only possible solution to the item. All other alternatives can clearly be judged as incorrect on the grounds of the fact that although *until*, *to*, as well as *till*, are also frequently used to refer to the end of a time period, they do not put a primary emphasis on this end point but rather on the duration of the period. *Until* and *till*, which are often used synonymously can, therefore, only be used "with durative verbs, *ie* verbs that denote a period of time" (Quirk et al. 1985: 690). Similarly, the use of *to* to refer to the end of a time period puts a particular emphasis on that period as such. It can be said that "in the temporal domain", *to* "can be used for the end point of a period of duration" (Huddleston & Pullum 2002: 660). The primary reference to duration expressed by *to* as well as *until* and *till* can further be proven by the fact that especially *to* but also the latter two prepositions are often used in combination with *from*. By denoting the starting as well as the end point of a particular time period, such constructions clearly emphasise duration. With reference to the distinction between durative and momentary semantic aspects, it can be said that especially *until* and *till* can be described as the direct counterpart of *by* (c.f. Quirk et al. 1985: 690).

From what has been just outlined it becomes possible to state that a primary focus of the item lies upon differences with regard to semantic aspects expressed by the

verb, such as momentariness as opposed to duration. The predicate of the sentence plays a crucial role in determining the choice of the correct preposition. It can, thus, be claimed that the context of the sentence is sufficient to justify alternative c as the only correct of the four alternatives provided for [5.4.2.2.I]. From a purely grammatical point of view, [5.4.2.2.I] can be considered to provide all of the information needed for successful processing of the item. However, as has been claimed above, the problem of presenting language in isolation from a broader context, which can in many cases appear rather artificial or unnatural, also applies to the testing of prepositions. As has been outlined before, there exists the danger that a rather restrictive testing format such as the multiple choice technique becomes even more restricted by the limitation of context. However, it can be claimed that [5.4.2.2.I], in contrast to the other items that are part of sections 2.1 and 2.2 of the TUC-PT, nevertheless includes a fairly high degree of authenticity. If one takes into account the fact that the TUC-PT is normally administered from the morning up to noon, the sentence forming the basis for [5.4.2.2.I], implies a relation to the actual real-world situation the testees are in at the moment they process the item or the test in general. The sentence could thus have been uttered in exactly the same form within an authentic communicative process that deals with the testing situation but that is not part of the test processing process itself. This relation to actual real-life communication that relates to an authentic situation every test taker experiences can be said to make the item somewhat more meaningful with a view to communicational aspects. However, it remains relatively restricted which is on the one hand due to the absence of a more extended context (either in the form of a dialogue structure mirroring a communicational process or a longer text) and on the other hand, to the multiple choice format in general which cannot be claimed to be very communicative (c.f. Heaton 1988: 11).

The second item in the TUC-PT that tests prepositions on the basis of the sentence is item three of section 2.2

[5.4.2.2.II]
My rabbit is bigger *like* / ***than*** / *as* my cat.

This item rather obviously focuses on the problem of adjective comparison which seems to be particularly difficult for German learners of English. Therefore, a number of aspects that are related to potential interference of the German mother tongue of most of the testees can be traced within the structure of the item. [5.4.2.2.II] yielded the fairly

low error ratio of 5% of all test takers. 1% and 4% of the target group decided for the first alternative, *like*, and the last alternative, *as*, respectively. Although the comparison of adjectives is clearly a problem area in German EFL contexts, it can be assumed from this that the target group has reached a level of English language (or grammar) competence at which such problems only play a rather insignificant role. The item will nevertheless be discussed here as there is only a limited number of items in sections 2.1 and 2.2 of the TUC-PT that can be used for a comparison of the testing of prepositions on sentence level as opposed to their testing on the basis of a coherent gapped text, such as section one of the WBPT. Furthermore, though not the primary topic of this paper, the problem of German-English-interference problems has already been touched upon several times and as [5.4.2.2.II] very clearly includes aspects related to this topic, its discussion might add an interesting component to the present analysis.

Before a brief discussion of the item itself can be provided, it is important to investigate the nature of the alternatives given for [5.4.2.2.II]. Although the item is part of a chapter that deals with the testing of prepositions it needs to be kept in mind that all of the three alternatives are "items which have functions that are difficult to classify in terms of traditional word classes" (Quirk et al. 1985: 661). All of the three words can for instance function as prepositions and conjunctions. Although it is sometimes arguable whether they function as preposition or conjunction in a respective case, a very broad definition of the difference between the two categories, which has already been referred to in this paper, will serve as a basis for clearly defining [5.4.2.2.II] as focusing on prepositions and not conjunctions. According to Quirk et al., this difference consists in the fact that "prepositions introduce complements which are nominal or nominalized, whereas the corresponding conjunctions (subordinators) introduce a subordinate clause" (1985: 660). On the basis of this definition, all three alternatives to [5.4.2.2.II] can be described as functioning as preposition as the complement, *my cat*, is an NP.

The basic challenge test takers are faced with by [5.4.2.2.II] is to recognize the fact that the sentence contains what Quirk et al. call a "COMPARISON OF NONEQUIVALENCE" (1128). This can for instance be inferred from the comparative form *bigger* included in the sentence. Such forms of comparison are typically expressed with the help of *than* in English (ibid.). The first reference to *than* cited in the *OED* is its use "after a comparative adjective or adverb [...] to introduce the second member of the comparison; [...] expressing the comparative of inequality" (s.v.), or "COMPARISON OF NONEQUIVALENCE" (Quirk et al. 1985: 1128).

The first distractor provided for [5.4.2.2.II], *like*, did not seem particularly attractive to the target group. This can be explained by the fact that because of their

English education at school, the testees are probably very well accustomed to the fact that *like* is not normally used for the comparison of adjectives in English. Rather, "*like* phrases commonly occur as complements to *be* or other complex-intransitive verbs, especially the appearance verbs *seem* and *appear*" (Huddleston & Pullum 2002: 1154). More interestingly, however, the comparison of inequality to be expressed in [5.4.2.2.II] with the help of *than* would in a German version of the sentence be expressed with the help of *als* which is the prototypical translation equivalent for English *as*. With reference to comparisons of inequality or the *Komparativ* in German, Sommerfeldt and Starke point out the following. "Bei Ungleichheit erscheint standardsprachlich stets *als* zur Einführung des Vergleichsmaßes" (1998: 133). As has been mentioned above, the most common translation of *as* is German *als*, in addition there exists a high amount of graphological and phonological similarity between the English and the German term. This can be assumed to be one of the reasons why the distractor *as* was chosen by most of the subjects who answered the item incorrectly. Ellis provides a definition of L1 interference when he describes the concept "as 'transfer'" or "'the influence that a learner's L1 exerts over the acquisition of an L2'" (1997: 51 qtd. in Bhela 1999: 23). On the basis of this it becomes possible to hypothesize that the incorrect selection of *as* as a solution to [5.4.2.2.II] could be a result of the test takers' projecting the particularities of German adjective comparison onto the English language. However, it cannot be proven without further analysis whether this was the case for the target group in question or whether other factors influenced test taker choice as well. Nevertheless, it can be assumed that the item focuses on potential interference problems and the testees' ability to avoid such problems as it provides a distractor which might under certain circumstances elicit a structure that is very similar to a comparative structure typical or most of the subjects' German L1 but which is not possible in English. However, as the analysis of error rates showed, the problem just mentioned did not seem to be a fundamental one for most of the testees.

As has been mentioned before, [5.4.2.2.II] can be described as a further example for sentence-based multiple choice gap filling items which are not rendered ambiguous because of the limited context that is provided. The adjective *bigger* contains sufficient information for testees with a suitable amount of competence with regard to the principles of adjective comparison in English to identify *than* as the only possible solution to the item. Nevertheless, the item also shows that gap filling tasks that are based on more or less isolated short sentences like the one that forms the context for [5.4.2.2.II] can appear relatively artificial and decontextualized when compared to real-life communicational situations. If presented in isolation from a meaningful context, a

sentence such as *My rabbit is bigger than my cat.* might appear very trivial to testees and it might at the same time evoke a picture of language in general and grammar in particular that does not seem to have a real-life equivalent.[53]

5.4.2.3 Summarising the Comparison

Summarizing what has been mentioned in the preceding sections it can be said that the presentation of multiple choice grammar test items within a relatively small amount of context does, in the case of the TUC-PT, not lead to ambiguities with regard to what are correct and incorrect solutions. Nevertheless, as the analysis of both the items that focus on tenses and those that focus on prepositions showed, the testing of these grammatical contents on the basis of isolated sentences can lead to rather artificial constructions, such as [5.4.2.2.II]. The analysis of section one of the WBPT clearly showed that the use of more or less authentic, coherent text materials as the basis for gap filling tests of grammar can restrict the spectrum of grammar points to be tested. At the same time, however, this practise has the advantage of presenting and testing grammatical structures in a form that is in any case closer to (written) real-life communication than the use of single sentences. It seems that for the testing of tenses the context of a broader text plays a more significant role as a source of information needed in order to identify correct or appropriate alternatives than for the testing of prepositions. This is probably due to the fact that temporal relations can be expressed in language in a variety of different ways. Besides the "grammatical categories of tense and aspect", "principles of discourse organisation, such as 'the order in which situations are reported corresponds to their temporal order in reality'" can for instance function as important indicators of temporality as well (Klein 1994: 14). Such aspects can be described to be inherent features of broader coherent texts which can serve as a very valuable basis for the decision for or against a particular tense form in grammar test items. Although for many of the tense items in section one of the WBPT, the context of the sentence or clauses immediately surrounding items contained sufficient information - particularly in the form of conjunctions and/or the tenses of other verb forms in the same sentence or clause - in order to identify the respective solutions, it can be claimed that the context of the texts as coherent entities formed a fundamental source of information as well. For the testing of prepositions which mainly focused on semantic and partly on syntactic properties of members of this word class, the context of the sentence or clause surrounding items functions as the most important cue. Although the respective base text in its entirety has an important supporting function for the

---

[53]With reference to this problem, c.f. Heaton (1988: 28; 36).

determination of the correct or appropriate solution to an item it can be claimed that this function is not as significant for the testing of prepositions as it is for the testing of tenses. Nevertheless, the previously mentioned significance of a broader, coherent, and if possible authentic language context can also be underscored with regard to the testing of prepositions.

Apart from the role of contextual features for the testing of particular areas of English grammar within different varieties of one and the same test method, the present paper is also concerned with the potential of different testing techniques as means of testing tenses and prepositions. The following sections will put particular emphasis on the question whether and to what extent the C-test format can be used as a grammar test instrument. For that purpose, individual items from section two of the WBPT will be analysed and described in closer detail. The findings of this analysis will then be compared with section 2.3 of the TUC-PT, a text-based multiple choice gap filling task. This is done with the primary intention to discuss the potential of the C-test as a measure of grammatical competence in comparison to more traditional grammar test methods. In order to have a solid basis for the discussion of C-test items in sections 5.5.1.1 and 5.5.1.2, the following chapter will focus on some theoretical aspects related to this technique.

5.5 The C-Test – General Assumptions

In the course of a growing research and teaching interest in integrative teaching and testing materials and techniques, the C-Test has gained considerable importance. In very general terms, the C-Test can be described as a variety of the cloze procedure. Raatz and Klein-Braley give a more concrete account of the method when stating the following:

> A C-Test is an integrative written test of general language proficiency based on the concept of reduced redundancy. A C-Test consists of five to six authentic texts, each complete as a sense unit in itself. In these texts the first sentence is left standing. Then the 'rule of two' is applied: beginning at word two in sentence two the second half of every second word is deleted. Numbers and proper names are usually left undamaged, but otherwise the deletion is entirely mechanical. (n.d.)

As can be seen from this definition, the C-Test is commonly described as a global measure of language proficiency in general as opposed to its individual components (c.f. Grotjahn 1995: 37;56). The idea of *reduced redundancy* mentioned in the definition provided by Raatz and Klein-Braley is based on the assumption of a close relation between a learner's competence in a foreign language and his/her ability to make use of the natural redundancy of this language (Grotjahn 1995: 38). As C-Tests

are normally designed according to mechanical principles, natural redundancies that are part of any written text are reduced by such deletion principles as the 'rule of two', the test is thus assumed to allow judgements about a testees general ability to process the language in question. [54] The relevance of C-Tests as measures of general language proficiency can for instance be seen if one takes into the account the presence of redundancy "in all levels of language", including "the lexicon, the semantics and the pragmatics of a language" (Raatz & Klein-Braley, n.d.)

C-Tests are nowadays frequently used as placement instruments at university level (Grotjahn 1995: 37). One of the reasons for this development is probably the fact that the C-test is a very economical means of language assessment, i.e. after a careful selection of texts has been carried out, it is fairly easy to develop, especially if one follows the classical deletion principle expressed by the 'rule of two'. Furthermore, C-tests are in general very easy to score as the number of possible alternative solutions for a given item is in most cases extremely restricted (Piper 1983: 48). Investigating the usability of the C-test for language testing at university level, Coleman provides an extended list of advantages of the format, including economy with regard to "time, staff-hours and money" as well as its high reliability and objectivity (1994: 218). Besides mentioning many of the obvious advantages, Coleman also refers to a number of drawbacks of the format, one of them being the fact that " it is unclear exactly what is being tested" (ibid.), i.e. C-tests, which are designed according to the general, mechanical principle mentioned above, focus heavily on testees' ability to process a language in its entirety which makes it relatively difficult to draw conclusions about their competence with regard to individual language skills or aspects of language proficiency such as grammatical competence. According to Piper it is nevertheless possible to use C-tests as a means of assessing particular aspects or "areas of a language", such as its "structure, semantic items, features of discourse, socio-cultural elements and many others" (1983: 47). This can be achieved via modification of the classical mechanical principle, i.e. by "selective deletion" which "offers the possibility of greater control over test items" (ibid.).

The fact that the WBPT uses a C-test in order to assess test takers' competence in the grammatical structure of the English language made it necessary to modify some of the typical characteristics of C-tests, including the use of a certain amount of such "selective deletion" (ibid.). Before some of the individual items of section two of the WBPT which focus on the use of tenses and prepositions will be analysed and discussed

---

[54] According to Piper, "the ability to allow for an incomplete or distorted message" as well as "the ability to recognize redundancy" are, amongst other things, essential skills "required in processing language" (1983: 46).

in closer detail, the C-test in the WBPT will be described with reference to the modifications of the general characteristics typical for the format that have been made.

5.5.1 C-Testing in the WBPT: Section Two - A Modified C-Test

Section two of the WBPT consists of a C-test with 25 items. As is the case for section one, there are four versions of section two stored in a database, each of them being based on a different text. Test takers are presented with one of these versions via a random principle. Examples of the individual versions can be found in AppendixIV. They are required to complete the items within the respective text by typing in the missing parts of words. For words with an even number of letters, half of these letters have been deleted, for words with an uneven number of letters the extra letter is left standing.[55] [56] For each item there is only one possible solution, thus, spelling mistakes etc. are judged as incorrect.

In contrast to the classical C-test, the items in section two of the WBPT were not developed by entirely mechanical ways of deletion. Rather, a number of items in section two were chosen deliberately in order to arrive at gaps in the text which can be assumed to put a main emphasis on grammatical contents. Nevertheless, it has to be maintained that not all deletions in section two were made in the way just outlined. As a fixed deletion rate is probably one of the most distinguishing characteristics of C-tests and as it allows a relatively well-balanced distribution of items within the test, section two was designed on the overall basis of deletion of every 4th word. This principle was, however, not strictly obeyed to. At many instances, gaps were chosen which were assumed to put a main emphasis on grammatical and/or textual competence as two of the components of language proficiency in the sense of Bachman (1990).

Besides employing a relatively large extent of selective deletion, the WBPT furthermore varies from classical C-test formats in that every test taker is provided with only one text. Normally, C-tests consist of several short texts of about 60 to 80 words' length which deal with different topics (Grotjahn 1995: 38). One of the most important purposes of this convention is to achieve a high degree of test fairness, which could not in all cases be guaranteed if a C-test is based on just one text dealing with a specialised topic (ibid.). However, the topics of all of the texts used in section two of the WBPT are very general in nature: Text one deals with the life and work of the author Jane Austen

---

[55]This is contrary to the traditional deletion practise in C-tests where the extra letter is normally deleted in such cases (c.f. Piper 1983: 48). Leaving the test taker with more letters than he/she has to reconstruct is, however, claimed to decrease the possibility for alternative answers (c.f. Grotjahn 1995: 43), which was the primary purpose for not deleting the additional letter in words with uneven numbers in section two of the WBPT.

[56]The task for section two of the WBPT does not specify the number of letters that have been deleted and have to be reconstructed.

and can be described as a fairly straightforward, informative text which does not use figurative language etc. and can thus be assumed to be easily understood by most of the subjects in the target group. Text two is a news text from CNN's on-line edition dealing with the preparation of the Olympic Games in Athens, a topic which most of the subjects can be presumed to have some knowledge of. Text three is an informative text form the BBC news web page, dealing with the United Kingdom, including the history, the economy and the population of the country. Although it is taken from an on-line news page, it cannot be described as a typical news text, but rather a very neutral collection of information. The topic of text four is the UK as well, however, the text puts a special emphasis on nature and natural phenomena, such as the weather. It was taken from the same source as text four, and can, like text two, be described as a genuine news text. The base texts for the different versions of section two are provided in Appendix V. Both, the topics of texts three and four, although the latter is somewhat more specific than the former, do not require test takers to possess particular factual knowledge which exceeds the amount of general world knowledge the majority of the subjects can be assumed to have. Summarizing, it becomes possible to suggest that the texts in section two of the WBPT are not biased in favour of some kind of topical knowledge, which would result in the need for several thematically distinct texts in order to be able to guarantee a suitable amount of test fairness.

It is often claimed that the materials used as a basis for C-tests should be authentic (Raatz & Klein-Braley, n.d.). As far as authenticity within the C-test section of the WBPT is concerned, texts two, three and four were taken from authentic sources and were shortened but only slightly adapted. Text one, however, originates from a textbook for learners of EFL at an upper-intermediate level (c.f. Appendix V). Raatz and Klein-Braley point out that it is under certain circumstances possible to use "material from equivalent text books" as the basis for a C-test (n.d.). In the case of the WBPT, text one proved particularly useful as it deals with a cultural topic which most of the subjects at 'Abitur'-level should be able to deal with.

Another difference between section two and classical C-tests, is the fact that the C-test in the WBPT uses shorter text material than is typical for traditional formats. According to Grotjahn, the texts in a classical C-test should consist of 60 to 80 words (1995: 38). If one takes into account the claim of Raatz and Klein-Braley that a C-test should contain between five and six such texts (n.d.), it becomes obvious that such tests employ text material which has a total length of at least 300 words. The four texts in

section two of the WBPT, however, have a mean length of only 186 words.[57] Whereas traditional C-tests "should have at least 100 items" (n.d.), the C-test in the WBPT consists of just one text with 25 gaps. The main reason for this modification is the fact that the WBPT has five different sections which all require a relatively high amount of reading. Therefore, presenting test persons with only one short text serves the primary purpose of easing the test taking process by reducing the amount of reading involved. Nevertheless, in further applications of the C-test for placement purposes, the use of more varied text materials and a higher number of test items could be useful as this can be assumed to increase the reliability and validity of such a test.

As has been mentioned before, not all text materials used in section two originate from authentic sources. In contrast to section one of the WBPT, where the highest mean score was achieved for the non-authentic text from the learners' textbook, the mean scores for section two reveal a different picture. Although one might suppose that the material taken from the learners' textbook might have been the easiest to process, the mean score for C-test three, which is based on an authentic text was highest, i.e. 88%. The mean score for C-test two which is based on an authentic news text was 82%. For C-test one, based on a non-authentic text, the figure is 80% and the lowest mean score, i.e. 75% was achieved by testees who were provided with C-test four, which again is based on an authentic news text. From these figures, it can be seen that there is considerable variation of mean scores, i.e. the four versions of section two cannot be regarded as equivalent test instruments as far as their difficulty level is concerned. There can be a variety of reasons for such differences, including the difficulty level of the input material, the selection of items, etc. Although a very interesting aspect, a detailed discussion of such possible reasons would exceed the scope of this paper and shall thus be omitted here. Despite differences with reference to mean scores in section two, it is interesting to note that all of the figures are considerably high, especially when compared to the results obtained in section one of the WBPT. The total mean score for all four C-test versions is 81% whereas that obtained for all gap filling texts in section one is 71%.[58] From this, it can be assumed that all the four C-tests in section two were easier for test takers to process than the more traditional multiple choice gap filling activities in section one. Whether this is a reason of difficulty of input material, of item selection, text length, and/or differences

---

[57] Text one consists of 197 words, for the text two the number is 196 words, and for texts three and four the figures are 163 and 189, respectively.

[58] For a detailed account of test taker scores for the different sections of the WBPT discussed in this paper, c.f. Appendix II.

between the two test methods in general or other factors, cannot definitely be said without further analysis.

Despite a number of restrictions which shall be discussed below, the C-test has the advantage that test takers have to directly produce an answer to the items. For the case of the WBPT this means that they cannot simply select an answer from a given list by clicking on it, they rather have to type in their solution. It, thus, becomes possible to investigate the actual mistakes different test takers make which might give interesting insights into their way of dealing with certain grammatical problems. The following discussion of section two will deal with those items focusing on competence in English verb tenses and prepositions, their main emphasis and the mistake patterns that could be identified. A special focus will be upon the usefulness of C-tests for testing grammar. In section 5.5.2, the results will then be compared with those for the gap filling activity on text basis employed in the TUC-PT trying to find similarities and differences with reference to the testing of English grammar across the two formats.

5.5.1.1 Testing Tenses

As shall be discussed in closer detail in the following sections, using the C-test format to test grammar might, under certain circumstances, restrict the tester's possibilities of assessing subjects' competencies in this particular field. This is first of all due to the fact that it is only a part of a particular word that test takers have to supply. Especially for the testing of verb tenses, which often consist of a combination of more than one word, this can result in items for which it is not exactly clear whether they test competence in the appropriate use of tenses within a given context or whether they simply focus on the ability to complete words which are part of a verb tense. Furthermore, the fact that for C-tests there are no alternatives testees have to choose from, it is impossible to draw conclusions about their competence with reference to contrasts between different tense forms.

In total section two contains 18 items that were constructed via the deletion of the second half of words that constitute or are part of an English verb tense. C-tests one and two contain 4 such items, respectively; for C-tests three and four the respective number of items is five. In contrast to section one of the WBPT, the distribution of these items is relatively even across the four individual C-tests. Nevertheless, the respective tenses that are tested within the four C-tests vary from text to text so that the four cannot be claimed to be fully equivalent measures of the same grammatical contents.

As has been mentioned earlier on, reconstructing tense forms or parts of tense forms in C-tests often seems to be a matter of competence in the formation of tenses,

i.e. the reconstruction of individual components of a tense form rather than their use. This can for instance be seen, if one takes a closer look at item eight of C-test version one.

[5.5.1.1.I]

Jane Austen couldn't (7) poss_____ (possibly) have **(8) imag_____** (imagined) this kind of worldwide (9) fa_____ (fame).

This item seemed to be problematic for more than one fifth of the test takers. 24% of a total of 21 subjects who were provided with C-test one answered it incorrectly. All of them did not seem to be able to recognize the need for a past participle form of the verb *to imagine*. In other words, all of the test persons who answered [5.5.1.1.I] incorrectly supplied the letters *i-n-e* which results in the formation of an infinitive form of the verb. The item can be described as part of the complex tense form *couldn't have imagined* which consists of a modal verb in combination with a form of *have* and a past participle. In order to be able to correctly complete the tense form, the test taker needs to recognize this structure which could only be completed suitably with a past participle which results in the formation of past form of the modal *could* in combination with the full verb *to imagine*. The seemingly high difficulty of the item can partly be explained by the complex structure of the verb phrase which additionally contains an adverb that is itself an item of the C-test. Being able to successfully complete [5.5.1.1.I] can thus be described as largely dependent upon test takers ability to understand the structure of the verb phrase the item is part of. However, the fact that the C-test provides large parts of the respective tense to be reconstructed, makes it questionable whether the item can be described as a measure of test takers' competence in the use of the past tense with complex verb structures containing modal verbs. Rather, finding the correct answer to [5.5.1.1.I] is largely dependent on competence in the formation of such constructions, i.e. the fact that they consist of a modal in combination with a form of *have* and a past participle of the full verb.

In contrast to [5.5.1.1.I], item 15 of C-test version one, does not primarily seem to test competence with regard to the components needed to form a particular verb tense.

[5.5.1.1.II]

..., and she **(15) ha_____** (had) earned a grand (16) tot_____ (total) of £ 648.65 from her (17) boo_____ (books).

This item was answered incorrectly by 24% of the total testing population provided with C-test one. Interestingly, all 24% of test takers supplied the letter *-s* instead of the letter *-d*, forming a simple present instead of the correct solution, a simple past form, of the verb *to have*. Although testees have to supply a simple past form of the verb *to have* in order to successfully complete [5.5.1.1.II], the item cannot be said to focus on the use of the simple past as opposed to other tense forms, such as the simple present. It is part of a compound tense, i.e. a past perfect tense, and it is this particular fact that test takers need to realize in order to find the correct solution to the item. As temporal relations are expressed by compound tenses such as the past perfect via the auxiliary and not the respective participle, [5.5.1.1.II] can be described as focusing on test takers' understanding of the temporal relations typically expressed by the past perfect as opposed to e.g. the present perfect. Most importantly, test takers need to recognize that the past participle directly following [5.5.1.1.II] is part of the same tense form as the item itself and that the choice of letters to complete the verb *to have* of which only the base is given, determines whether a present perfect or a past perfect tense is formed. Finding a correct solution for [5.5.1.1.II] can thus be described as focusing to some extent on the formation of compound tense forms.

Much more interestingly, however, the central problem posed by the item is that of deciding whether a past or a present perfect tense is more appropriate in the context in question. In order to make this decision, test takers need to prove an understanding of the temporal relations expressed by the context of the item. For [5.5.1.1.II] it is first of all the context provided by the sentence the item is part of that can serve as a basis for finding an appropriate solution. The full sentence is given below.

[5.5.1.1.III]

When she died spinster (12) i_____ (in) 1817, only four of (13) he_____ (her) novels had (14) be_____ (been) published, all anonymously, and she **(15) ha_____** (had) earned a grand (16) tot_____ (total) of £ 648.65 from her (17) boo_____ (books).

Clearly, the sentence makes reference to a definite time point in the past, i.e. the death of Austen, which is explicitly indicated by the simple past verb form *she died* in

combination with the PP *in 1817*. All the other situations and/or events expressed by the VPs in the following clauses are doubtlessly identifiable as having occurred before this time in the past. [5.5.1.1.II] is part of one of these VPs which shares one of the most crucial characteristic of the past perfect, i.e. it is "context dependent" or, more precisely, "referring to a time before a past time signalled elsewhere" (Biber et al. 1999: 469). Therefore, the choice of the letter *s* in order to complete [5.5.1.1.II] can doubtlessly judged as an incorrect answer to the item as the present perfect cannot express such temporal relations. Contrary to the past perfect, it does not relate situations or events in the past to another, later, time in the past but "to a present time orientation" (Quirk et al. 1985: 192).

The definite time point expressed by the simple past form in the beginning of the sentence, [5.5.1.1.II] is part of, thus clearly indicates the need for a past perfect tense. Furthermore, the VP in the clause preceding [5.5.1.1.II] is in a past participle tense and although a part of this VP, the past participle of the verb *to be* which indicates passive voice, is itself an item of C-test one, the construction can relatively easily be identified as a past participle as the most important components needed to do so, namely the auxiliary *had* and the past participle are given. This might serve as a further cue in deciding for the correct solution to [5.5.1.1.II].

Both, the VPs *had been published* and *had earned* fulfil the same function in the text, they refer to events/situations that took place before a definite time in the past. [5.5.1.1.II] as well as item fourteen of C-test one can therefore be described as not only focusing on the use of the past perfect tense in isolation from the context within which it occurs but also on test takers' ability to recognize temporal relations between clauses. [59]

For [5.5.1.1.II] many test takers were not able to produce a solution that results in the formation of a correct verb tense. For item eleven of C-test version two, however, recognizing that the item is part of a verb tense seems to have been the most important problem.

> [5.5.1.1.IV]
>
> (8) Fre_____ (French) engineer Pierre Bideau, (9) wh_____ (who) designed the millennium lights-and-fireworks spectacle (10) a_____ (at) the Eiffel Tower, **(11) i_____** (is) leading the Athens (12) proj_____ (project).

The item was answered incorrectly by 31% of a total testing population of 26 persons that were presented with C-test two. All of the test persons, who provided an incorrect solution to [5.5.1.1.IV], did not seem to be able to recognize the need for a finite verb form in the clause the item is part of. Seven out of eight, i.e. 27% of the total group of testees for C-test one, supplied the letter *n* which results in the formation of the preposition *in*.[60] Together with the *-ing* form *leading* which follows the item, this can be described as resulting in a construction that is similar in form to a combination of preposition and gerund or an *-ing* participle. Such a construction is, however, grammatically inappropriate in the context of [5.5.1.1.IV] because *-ing* forms, such as *-ing* participles, cannot be the used in clauses such as the one the item is part of (c.f. Quirk et al. 1985: 150).

The fact that so many test takers were not able to realize the need for a verb form within the clause [5.5.1.1.IV] is part of, implies that for this group of subjects, the item did in fact not seem to focus on competence in the use of verb tenses but rather on syntactic competencies. Precisely, the verb form *is leading*, which results out of the correct solution to [5.5.1.1.IV], is part of the clause *French engineer Pierre Bideau, [...] is leading the Athens project*. This is a very common clause type in English, comparable to a simple sentence following an *SVO* pattern. Understanding the pattern of this clause should automatically lead test takers to recognize the need for a finite verb form in the clause as "the V element in a simple sentence is always a finite verb phrase" (Greenbaum et al. 1990: 204). The fact that such a high proportion of the testing population were not able to realize this, can probably be explained by the non-defining relative clause , *who designed the millennium lights-and-fireworks spectacle the Eiffel Tower,* being embedded in the clause. Problems with the syntactic structure of the sentence are probably further intensified by the fact that the embedded relative clause contains itself two further C-test items.

Taking into account the important role that the syntactic structure of the complex sentence, [5.5.1.1.IV] is part of, it becomes possible to assume that the item does not primarily focus on test takers' competence in use of the present progressive tense but rather on other aspects of English grammar. If the need for a finite verb is recognized by

---

[60]One test taker (4% of the total testing population) supplied the letter *-t* which results in the formation of the personal pronoun *it*. As only one person cannot be regarded as a representative sample of test persons and because it is relatively probable that the choice of the letter *t* is due to a typing mistake, this solution shall not be taken into further consideration in this paper.

the test taker, it can be assumed that the verb form *is* which results out of the provision of the correct answer to [5.5.1.1.IV] is a rather obvious way of completing the structure in question. It can thus be assumed that [5.5.1.1.IV] too obviously evokes the letter *s* as an appropriate solution, once the need for a finite verb form has been realised. Supplying the correct answer to the item automatically results in a present progressive verb tense. Ideally, test takers should nevertheless verify against their competencies in the use of tenses whether the so formed tense is appropriate in the context provided by the text the item is part of. It is, however, not clear whether [5.5.1.1.IV] assesses competence in the use of the present progressive tense to a suitable extent as the proportion of the construction that has to be reconstructed by the testee is fairly low. It needs to be maintained that a direct assessment of tense-related competencies would probably have been easier to achieve via a multiple choice item which presents the correct present progressive form *is leading* along with two or more alternative tense forms of the verb. Such an item would explicitly ask test takers to consciously select this particular tense and to compare its properties and uses with those of other tenses. In contrast to multiple choice items focusing on tenses which normally provide only verb forms as alternatives, item [5.5.1.1.IV] directly demands test takers to recognize the need for a finite verb form within the clause structure in question. Recognizing certain properties of the context which justify the use of a present progressive form of the finite verb can be assumed not to be the most crucial criterion for providing the correct answer to the items. Such context related information can nevertheless be found in the text that forms the basis for C-test two. One of the main topics of the text is "a project to highlights the [...] monuments" of the city of Athens "during the Olympics" ("Ancient Athens I", Appendix V). In the first part of the text, one part of this project is more closely described following this description the text explains the future outlook of the project (c.f. ibid.). The context of text two thus clearly implies that the project is under progress, i.e. not finished at the TU. Accordingly, also the function of Pierre Bideau as leader of the project can assumed to be in progress at the time the text was written. With reference to the progressive aspect, Quirk et al. cite the following "three components" that are or can be typical for this type of aspect. Verb phrases that have a progressive aspect can indicate that a "happening has DURATION", that it "has LIMITED duration", and/or that "the happening is NOT NECESSARILY COMPLETE" (Quirk et al. 1985: 198). With reference to the context provided by text two it becomes relatively obvious that the meaning of the verb phrase [5.5.1.1.IV] is part of implies a certain incompleteness of Bideau's leadership of the project. Another, very crucial component of the progressive aspect especially in combination with the present tense is the fact that

it indicates a certain temporariness (Leech & Svartvik 1994: 67). This characteristic of the present progressive which can be compared to what Quirk et al. label "LIMITED duration" (1985: 198) is also inherent in the context provided by text two. This can be exemplified by the fact that the project the article deals with is concerned with the Olympic Games in Athens which evidently only have a limited duration to which also the project must be restricted. Therefore it can be noted that for those test takers who provided the correct solution to [5.5.1.1.IV], the context surrounding the item could easily serve as a means for a conscious reflection whether their solution is appropriate. However, although typical contextual characteristics that call for the use of a present progressive form are implied in the context, test takers' competence with reference to such characteristics or the use of the present progressive cannot definitely be proven to be assessed by [5.5.1.1.IV]. Although it is to a certain extent possible to adjust the C-test format via "selective deletion" in such a way that a special focus on certain areas of language proficiency, such as grammatical competence becomes possible (c.f. Piper 1983: 47), the analysis of [5.5.1.1.IV] very clearly implies that it is at some instances rather difficult to identify which particular grammar point is tested with which degree of explicitness – or, as Coleman quite drastically puts it, " it is unclear exactly what is being tested" (1994: 218). Although this statement was originally made with reference to classical C-test formats which focus on the assessment of language proficiency in general (c.f. ibid.) it can also be applied to testing contexts where the C-test is used to assess particular proficiency components.

Like [5.5.1.1.IV], item eleven of C-test version three is also concerned with the present progressive tense.

[5.5.1.1.V]

The economy – (6) on_____ (one) of the (7) larg_____ (largest) in the world – (8) i_____ (is) no (9) lon_____ (longer) manufacturing- but services-based, (10) an_____ (and) e-commerce is **(11) beco_____** (becoming) more and more important.

Whereas for [5.5.1.1.IV] test takers had to provide the second half of the form of *be* as one part of a present progressive verb form, for [5.5.1.1.V] they need to reconstruct the missing letters of an *-ing* form accompanying the form of *be* in such a construction. When compared to those for [5.5.1.1.IV], error rates for [5.5.1.1.V] were considerably lower. 10% of a total testing population of 30 for C-test three answered [5.5.1.1.V]

incorrectly. Within this group of testees there is no variation with reference to the incorrect letter combinations that were supplied, i.e. all 10% of test persons completed the item with the letters *m-e* which results in the formation of an infinitive form of *to become* rather than an *-ing* form which would have been the correct solution.

In contrast to [5.5.1.1.IV], it is not primarily recognizing the need for a finite verb which seems to be crucial for finding the correct solution to [5.5.1.1.V]. The finite verb, i.e. the form *is*, is already supplied by the context. Rather, the item requires test takers to identify the close relation that exists between *is* and the item itself, which together form a present progressive verb tense. Taking into account the recurring mistake pattern for [5.5.1.1.V], it seems reasonable to suggest that for a relatively high number of subjects this relation was not evident. The context of the item, however, clearly implies a process that is in progress which rather explicitly calls for a verb form containing the progressive aspect (Quirk et al. 1985: 197). The context of text three refers to processes of economic change in the United Kingdom (c.f. Appendix V, "Country Profile United Kingdom I"). The reference to a certain "happening in PROGRESS" (Quirk et al. 1985: 197) which needs to be expressed by the verb phrase [5.5.1.1.V] is part of is further emphasised by the expression *more and more important* which obviously bears implications of the progression of a certain state or situation as well as the fact that this state or situation has not yet reached an end. The type of meaning to be expressed by the VP the item is part of is, thus, comparable to what Quirk et al. label "EVENT PROGRESSIVE", i.e. "it conveys the idea that an event has duration, and has not yet come to an end" (199). Therefore, the need for an *-ing* form of the verb *to become* in [5.5.1.1.V] can very clearly be seen from the context of the item.

It needs to be emphasised, however, that is not possible to explain the mistake pattern for [5.5.1.1.V] mentioned above solely on the basis of insufficient knowledge of conventions for the use of the present progressive. It also needs to be taken into account that the test takers' inability to recognize the close relation that exists between the item and the verb form preceding it could be due to general problems with the testing format of the C-test. According to Weir, "the technique suffers from the fact that it is irritating for students to process heavily mutilated texts and the face validity of the procedure is low" (49).[61] A number of test takers reported certain problems with the section two of the WBPT, the most frequent of them being unfamiliarity with the C-test format. It could, thus, also be the case that the format itself prevented test takers from recognising tense constructions which they would otherwise, e.g. within another test format, have

---

[61] The term *face validity* refers to "the degree to which a test appears to measure the knowledge or **abilities** it claims to measure, as judged by an untrained observer", such as the testee him/herself (Davies et al. 1990: 59).

been able to identify. The question whether this is the case or not cannot satisfactorily be answered within the scope of this paper. However, it is certainly an interesting question and would constitute a good starting point for an analysis of C-testing as opposed to other testing techniques from a perspective that is related to processes of test taking. In the following section of this paper, the potential of the C-test as an instrument for the testing of English prepositions will be discussed and analysed.

5.5.1.2 Testing Prepositions

As was the case for the testing of tenses with the help of the C-test format, the following section will, amongst other aspects, raise the question whether the method can doubtlessly be described as a measure of competence in the use of English prepositions or rather as a mere instrument for the testing of knowledge of their lexical forms. In total, section two of the WBPT contains 16 items focusing on prepositions. Whereas the distribution of tense items across the four texts used in section one is relatively even, it needs to be pointed out that this is not the case for the 16 preposition items. While text one contains three items focusing on prepositions, for text two, the figure is as high as nine and for texts three and four the number of items is two, respectively. Accordingly, the four texts cannot be understood as equivalent test instruments which is not only a reason of the different materials used but also of the number of items focusing on particular aspects of grammatical competence. It is thus highly important for further applications of the test to be valid means of assessment to increase equivalence between the four versions of the C-test in section two. This could be achieved either via the selection of texts and deletions so that all of the four versions contain approximately the same number of preposition items. As has been mentioned earlier on, the C-test in the WBPT was designed with a main focus on grammatical problems, nevertheless a certain amount of mechanical deletion, which is one of the major characteristics of the C-test format, was maintained. It can be hypothesised that it is extremely difficult to design four different versions of a C-test which follow a basic mechanical deletion principle with a number of modifications; this is basically due to the fact that a high number of text materials, which contain grammatical patterns of a very similar nature, are relatively difficult to find. If one nevertheless attempts to use C-test texts for which the number of adaptations is kept as small as possible, it could therefore be useful to reduce to number of alternative C-test versions in section two and thus provide fewer but more equivalent versions with regard to the grammatical contents they test.

As has been claimed in section 5.5.1.1 not all items in the C-test can be viewed as focusing primarily on the use of a particular grammatical pattern, in this case tenses. As shall be discussed in closer detail in the course of this chapter, for some of the items in section two that focus on prepositions, a similar picture emerges. However, one item which can be described to focus relatively directly on the use of the preposition *of* in English is item 19 of C-test version one.

[5.5.1.2.I]

sales **(19) o**_____ (of) her novels rival (20) mod_____ (modern) bestsellers.

[5.5.1.2.I] was answered incorrectly by 29% of the total testing population of 21 for C-test one. Almost half of the testees who were not able to find the correct solution, *of,* for [5.5.1.2.I] supplied the letters *-v-e-r* which results in the formation of the preposition *over*, another relatively common mistake on the item was the choice of the letter *-n* to fill the gap which leads to the preposition *on*; this mistake was made by 10% of the total testing population. Obviously, those subjects who did not identify the letter *-f* as the correct solution to [5.5.1.2.I], were not able to recognise the central problem the item is corned with. It primarily deals with the use of *of* as a grammaticised preposition expressing a genitive-like relation between the NPs *sales* and *her novels,* the latter being the prepositional complement of *of*. In the construction under focus, *of* has the function of relating "two NPs, one forming the a constituent of the other" (Huddleston 1984: 268). Such relations are typically expressed in English with the help of the genitive, or as in the case of [5.5.1.2.I] an *of*-construction.[62] Expressing relations "between two entities, one being that represented by the prepositional complement" (Quirk et al. 1985: 673), is not only a typical characteristic of *of* used in constructions like the one to be discussed here. In general terms, this function can be said to be typical for all prepositions (ibid.). From a semantic point of view, the most typical of such relations can probably be described as one related to space or physical location, or, as Huddleston and Pullum put it, "most of the central prepositions in English [...] have meanings that are quite clearly locational in origin" (2002: 647). Taking a look at the most common incorrect solutions to the item, it becomes obvious that almost all of the subjects who supplied an incorrect solution chose a preposition which has or can have locational meanings. The primary sense of *over* is according to the *Longman Dictionary of the English Language (LDEL)* "across a barrier" (s.v.). For *on*, the incorrect solution to

---

[62]For a detailed account of differences in the use of the genitive and the of-phrase, c.f. e.g. Greenbaum (1996: 112-114).

[5.5.1.2.I] which was chosen with the second highest frequency, the use cited first in the *LDEL* is that which indicates "support from below" (s.v.) Therefore, it is possible to suggest that the subjects just mentioned did not recognise the need for a preposition that does not have locational (or temporal) meanings but rather "no identifiable meaning independent of the grammatical construction in which it occurs" (Huddleston & Pullum 2002: 601). It can therefore be assumed that [5.5.1.2.I] does not primarily focus on the semantics of the preposition *of* but rather on its use in a construction that expresses what Leech and Svartvik call a "**verb-object relation**" (1994: 61), namely the fact that the novels were sold.

It can furthermore be noted that the clause structure provided by the immediate context of the item might serve as an important cue for testees to find the correct solution. It is typical for the grammaticised preposition *of* to act "as a post modifier in noun phrases"(Quirk et al. 1985: 703). In *of*-constructions, a "superordinate noun phrase precedes a noun phrase introduced by *of*" (1276). Thus, in order to solve [5.5.1.2.I] correctly, it is crucial for the test taker to understand that certain prepositions can have "specific syntactic roles in the language that are not determined by their meanings" (Huddleston & Pullum 2002: 647). Furthermore, the item focuses on test takers' being able to recognise the specific relation that exists between the NP *sales* and the NP *her novels* which is clearly genitive-like and does not bear any implications of locational or temporal meanings, which are typical for many members of the word class of prepositions in English.

Although grammaticised uses are not the most prototypical of English prepositions, *of* is nevertheless "the most common preposition" in the language (Quirk et al. 1985: 703); it occurs almost exclusively in grammaticised uses. The target group can, thus, relatively safely be assumed to be familiar with *of*-constructions of the type focused on by [5.5.1.2.I]. Whether mistakes made on the item are due to insufficient competence in the use of the preposition *of*, or to other factors such as problems with the meanings of the words *sales* and *novels* which might have prevented testees from recognising the objective genitive structure contained in the clause in question cannot be found out without further investigation. Nevertheless, it is possible to summarise that although [5.5.1.2.I] focuses on a very common grammatical pattern, the item seemed problematic for a fairly high number of test persons. Another reason for such problems could be the fact that the C-test format only provides the initial letter of the preposition to be completed. The fact that in English there exist a relatively high number of prepositions starting in *o* provides a very high amount of alternatives for completing the item. [5.5.1.2.I] can thus be described as imposing less restrictions on testees' choice of

a particular solution than for instance a multiple choice item which presents the correct solution *of* along with two or more distractors. It can only be hypothesised whether this relatively high freedom with regard to the production of a grammatically correct solution to a problem such as the one presented by [5.5.1.2.I] can be regarded as a possible source for the relatively high error rate.

Generally speaking, items in section two of the WBPT that focus on the grammaticised use of *of* were a very frequent source of error. Another item which clearly focuses on an objective genitive-like relation expressed by the preposition *of* is item 16 of C-test version four.

[5.5.1.2.II]

The (13) proj_____ (project) asks people (14) acr_____ (across) the UK to (15) rec_____ (record) the timing **(16) o_____** (of) natural events

---

This item was answered incorrectly by 18% of a total testing population of 17 test persons who were provided with C-test two. 12% of the total target group for C-test two supplied the letter -*n* which led to the formation of the preposition *on*. As has been mentioned before, one of the most common senses of *on* in English is locational in nature. This makes it possible to suggest that those subjects who decided for *on* as the solution to [5.5.1.2.II] were not able to recognise the need for a grammaticised preposition in order to express an objective genitive relation. Summarizing, it can be assumed that [5.5.1.2.I] as well as [5.5.1.2.II] very clearly focus on the use of prepositions in specific syntactic constructions and, more importantly, on uses of a preposition which cannot directly be related to the semantic aspects but rather to the grammatical functions of this preposition within clause structures.

The same can be said to be true for another item focusing on the use of *of*. As was the case for [5.5.1.2.I] and [5.5.1.2.II], for item three of C-test version two, test takers need to supply the letter *f* in order arrive at the correct solution *of*.

[5.5.1.2.III]

The (1) lig_____ (lights) around the (2) oth_____ (other) high points **(3) o_____** (of) the city (4) we_____ (were) dimmed

---

Although error rates on this item were lower than those for [5.5.1.2.I] or [5.5.1.2.II], it still seemed to be problematic for a relatively high number of test takers. [5.5.1.2.III]

was answered incorrectly by 15% of the 26 subjects who were provided with C-test two. The majority of these 15%, i.e. 11% of the total testing population, supplied the letters *v-e-r* which led to the formation of *over*. Only one test person (4% of the total number of test takers) decided for the letter *n* and thus the preposition *on* as a solution for [5.5.1.2.III] Interestingly, *over* as well as *on* were also relatively frequent mistakes for [5.5.1.2.I] and [5.5.1.2.II]. As was the case for the latter two items, all of the test takers who provided an incorrect answer decided for a letter or combination of letters which led to the formation of prepositions that are commonly associated with locational meanings. This suggests that for [5.5.1.2.III] as well as the two items discussed before, those test takers who provided an incorrect solution were not able to recognise the central problem the item deals with, namely the grammaticised use of prepositions. As was the case for [5.5.1.2.I] and [5.5.1.2.II], for [5.5.1.2.III] the correct solution *of* has the function of expressing a genitive-like relationship between two NPs, here *high points* and *the city*. In contrast to [5.5.1.2.I] and [5.5.1.2.II], however, the kind of relation between the two NPs is not objective, but one expressing possession, i.e. the fact that the high points are part of or belong to the city. Greenbaum gives a definition of possessive meaning which can be expressed by the genitive as well as the *of*-phrase: "In a liberal interpretation, we could count as possession any connections between the two nouns [or NPs; K.U.] where the verbs *possess* or *have* can be used in a paraphrase" (1996: 113f.). Clearly the construction to be discussed here can be paraphrased in such a way, e.g. as *the high points that the city has*.

As has been mentioned before, it is interesting to note that for all of the three items that focus on the use of grammaticised *of*, similar mistake patterns can be observed. In the following, a brief discussion of the most frequent incorrect or inappropriate solution to [5.5.1.2.III], *over*, will be provided. Such a discussion is useful because the WBPT claims to be an objective test instrument and it is thus important for the solution(s) to an item specified in the answer key to be "genuinely correct" (Alderson et al. 1995: 47).

The preposition *over* can be identified as an incorrect or inappropriate solution to [5.5.1.2.III] on the grounds of the following assumptions. The context provided by text two, especially by its introductory sentence, clearly implies that the high points mentioned in the text are a part of the city of Athens which needs to be expressed with the help of the preposition *of*. The first sentence of text two says

[5.5.1.2.IV]
In Athens, the Parthenon and other ancient moments on the Acropolis Hill were bathed in honey-colored light.

Clearly, the Acropolis Hill is one of the high points of Athens which becomes especially obvious even for test takers who are not familiar with this fact if one takes a look at the term *Hill*. The sentence forming the immediate context of [5.5.1.2.III] establishes a cataphoric reference to the Acropolis Hill which, was mentioned in sentence one of the text, when it says

> [5.5.1.2.V]
>
> The (1) lig_____ (lights) around the (2) oth_____ (other) high points (3) o_____ (of) the city (4) we_____ (were) dimmed as (5) hun_____ (hundreds) of spotlights (6) aro_____ (around) the Acropolis were (7) swit_____ (switched) on.

It becomes especially evident from the expression *the other high points* that the paragraph in question deals with locations that are, like the Acropolis Hill, a part of the city of Athens. Therefore, the possessive relation between *high points* and *the city* can be reconstructed on the basis of cohesive processes within the text. The fact that both the Acropolis Hill as well as the other high points of Athens dealt with in the paragraph in question are referred to in close relation to each other implies a certain degree of redundancy which hints to the fact of C-tests being based on principles that try to make use of the natural redundancy of language during the testing process. Therefore, it can be assumed that the context of text two and the cohesive processes that are part of this context more or less explicitly determine *of* as the correct solution to [5.5.1.2.III].

The preposition *over* can, however, not only be determined as unsuitable in the context of [5.5.1.2.III] on the grounds of the aspects just mentioned but also simply on the basis of its possible meanings. According to *The New Oxford Dictionary of English* (*NODE*) one of the primary senses of *over* is "extending directly upwards from" (s.v.). Furthermore, it needs to be underscored that "*over* is prototypically a preposition of path rather than place" (Lindstromberg 1997: 113). Thus, the semantics of the preposition *over* can relatively clearly serve as a basis for identifying *over* as incorrect in the context in question.

Taking into account the facts just mentioned it becomes possible to suggest that the relation between the NP *high points* and the NP *the city* is clearly one of possession. On the basis of the context as well as common world knowledge, the high points can be identified as a part of the city and not as objects that bear any relation to an upward path

113

as generally expressed with the help of *over*. Therefore, *over*, can doubtlessly be identified as an incorrect solution to [5.5.1.2.III].

Whereas the three items just discussed focused on the grammaticised use of prepositions, item ten of version two puts a special emphasis upon the semantics of English prepositions, precisely the preposition *at* in a locational sense.

> [5.5.1.2.VI]
>
> (8) Fre_____ (French) engineer Pierre Bideau, (9) wh_____ (who) designed the millennium lights-and-fireworks spectacle **(10) a_____** (at) the Eiffel Tower

[5.5.1.2.VI] was answered incorrectly by 50% of a total testing population of 26 subjects. Taking a look at the actual mistakes that were made on the item, it becomes notable that problems with [5.5.1.2.VI] seem to be due to semantic as well as syntactic aspects related to the use of prepositions.

As far as the semantics of the correct solution *at* are concerned, it is obvious that the item calls for *at* in its locational meaning. According to the *Oxford English Dictionary*, this is the primary sense of the preposition (c.f. s.v.). With this meaning, *at* refers to "local position; answering the question *Where?*"; it thus expresses "the most general determination of simple localization in space", or "the simple relation of a thing to a point in space" (s.v.). Apart from providing a general account of the primary local sense of *at*, this definition also contains a reference to one of the most important and at the same time most basic characteristic of prepositions in general, namely the fact that they express "a relation between two entities" (Quirk et al. 1985: 657). Assuming *at* as the correct solution to [5.5.1.2.VI] it seems that the author of text two intended to refer to a very general local relation between the NPs *the millennium lights-and-fireworks spectacle* and *the Eiffel Tower*. Interestingly, however, 15% of the total target group that was provided with C-test two supplied the letters *-r-o-u-n-d* which results in the preposition *around*. Whereas it was possible for [5.5.1.2.III] to identify the incorrect solution *over* as definitely incorrect on the basis of contextual features as well as the semantics of the preposition itself, this seems somewhat more complicated in the case of *around* in the context of [5.5.1.2.VI]. The meaning of *around* is clearly more concrete than that of *at*. The former basically refers to positions that are situated "on all sides of" or "in all directions from" a particular object, person, etc. (*OED*: s.v.), however, the word can also "have a vaguer meaning of 'in the area of'" (Quirk et al.

1985: 681). The clause [5.5.1.2.VI] is part of is a non-defining relative clause which serves the primary function of further characterising the French engineer Pierre Bideau, who plays an important role for the project the text deals with (c.f. Appendix V, "Ancient Athens I", "Ancient Athens II"). The millennium lights-and-fireworks spectacle the relative clause refers to as well as its location can thus be assumed not to play a central role for the topic of the text in its entirety. It therefore seems probable that the reference to the location of the spectacle is a rather general one which would best be expressed by the use of *at* instead of *around*. Nevertheless, taking into account the possible uses of *around* mentioned above, it seems that this preposition cannot be judged as definitely incorrect in the context in question. This hints to the problem of how to deal with alternative solutions in C-testing (c.f. Raatz & Klein-Braley, n.d.). Especially for computer-based tests such as the WBPT that are intended to be scored objectively it is highly important to either design items in such a way that there are no alternative solutions possible or to include all possible solutions in the answer key. Finding all of the acceptable solutions to an item can in practise be a relatively complicated undertaking which requires a number of steps such as more or less extensive pre-testing or the discussion of items with different experienced language teachers and/or testers as well as native-speakers of English. Although alternative solutions to C-test items are a rare phenomenon (c.f. ibid.), i.e. much less frequent than for instance in cloze tests, [5.5.1.2.VI] shows that there exists the possibility of test takers providing a solution to an item which is acceptable grammatically as well as lexically, though not identical with the original version that is stored in the answer key. Therefore, it is important for further applications of the WBPT to revise [5.5.1.2.VI] in such a way that alternative solutions are either avoided completely or to include the letter combination *-r-o-u-n-d* in the answer key although the semantic implications expressed by the choice of this solution differ to a certain extent from those expressed by the preposition *at*. Another way of avoiding problems such as the one just mentioned would be to revise the task specification for the C-test so that it explicitly points out the number of letters to be supplied in relation to the number of letters provided. In its present version, the task specification for the C-test explicitly states that for some of the items only one letter has to be supplied, however, test takers are not directly referred to the fact that the number of letters to be supplied for each gap equals the number of letters that are provided or is at some instances one lower than the number of letters given.

[5.5.1.2.VI] focuses on the semantics of the preposition *at* as opposed to other members of the word class and thus on lexical competencies which are following

Bachman a component of grammatical competence (c.f. 1990: 87). Apart from this the syntactic structure of the clause the item is part of clearly forbids a number of word classes as solutions to [5.5.1.2.VI]. Generally speaking, the item together with the NP following it, act as postmodifier of the phrase *lights-and-fireworks spectacle* which is the head of a complex NP. Post modification within complex NPs is, according to Quirk et al. frequently achieved with the help of prepositional phrases, non-finite clauses, relative clauses, or complementation with a comp-element (1985: 1239; 1130). This implies that supplying letters or combinations of letters that result in the formation of a finite verb or other word classes such as conjunctions etc. will not lead to a grammatically correct syntactic structure.

Nevertheless, 4% of the total testing population supplied the letters *-r-e* for [5.5.1.2.VI] creating the finite verb form *are* which is syntactically and semantically impossible in the present context. As many as 23 % of all test takers for C-test two provided the letter *-s* as a solution to the item, which results in the formation of *as*. Since *as* can have the function of a preposition, this can under certain circumstances be judged as syntactically acceptable, however, semantically the word is not suitable because a certain local relation needs to be expressed. The same can be said about the solution provided by another 8% of the total number of test persons. This group of people supplied the letters *-n-d*, forming *and*. The syntactic structure of the clause would theoretically permit the use of *and*. However, *and* is generally used for purposes of coordination which implies that "two or more units of the same status on the grammatical hierarchy may constitute a single unit of the same kind" (Quirk et al. 1985: 46). On the grounds of common factual knowledge it should thus be obvious for test takers that the NPs *the millennium lights-and-fireworks spectacle* and *the Eiffel Tower* do not have the same status, i.e. they were not both designed by the subject of the clause, Pierre Bideau.

The relatively high number of different incorrect solutions provided by the target group for [5.5.1.2.VI] shows that C-tests provide test takers with a fairly high amount of freedom as far as the choice of a suitable solution to an item is concerned. This is especially true for cases such as [5.5.1.2.VI] where only a very small proportion of the word to be reconstructed is provided. As there exist a high number of prepositions in English which consist of only two letters, the C-test format seems quite useful for the testing of this particular grammatical category. However, this judgement about the usefulness of the C-test for the testing of prepositions needs to be further qualified. Although it seems definitely possible to assess test takers' competence in the formation and use of prepositions in particular contexts in a quite useful way, this first of all seems

to be true for simple prepositions that consist of a small number of letters, preferably two letters. Furthermore, in order not to provide testees with too much information about the preposition to be formed on a particular item, it seems that the format is most applicable for the testing of such prepositions that start in letters which can be identified as the initial letter of a certain number of prepositions and/or members of other word classes. As could be seen from the analysis of [5.5.1.2.VI], the C-test format does not restrict test takers' choice of a solution to a particular word class whereas many forms of traditional multiple choice gap filling techniques frequently provide members of the same word class as alternatives for an item. Whether this characteristic of the C-test is judged to be an advantage or a disadvantage is certainly dependent upon the respective purpose of a given test. In a situation where testers are most of all interested in test takers' ability to distinguish between the use of one member of a particular word class, there will probably be no particular interest in the opposition of this word class and others. Nevertheless, as could be seen from [5.5.1.2.VI], the C-test can be a very useful tool for determining whether a testee has sufficient competence in the use of members of one word class, such as prepositions, as opposed to others in particular syntactic and semantic contexts.

As has been mentioned before, the potential of the C-test to focus on the aspects mentioned above is to a high extent dependent on formal characteristics of the words to be reconstructed. Whereas for [5.5.1.2.VI] a fairly high amount of freedom with respect to the reconstruction of the missing part of the word in question was notable, for item 17 of C-test version two the situation seems to be a different one.

[5.5.1.2.VII]

Athens is keen (15) t_____ (to) present its (16) anci_____ (ancient) heritage **(17) dur**_____ (during) the Olympics.

[5.5.1.2.VII] was answered incorrectly by none of the subjects analysed for the purposes of this paper. Therefore, it can be presumed that the potential of [5.5.1.2.VII] to discriminate between testees' competencies is rather low. According to Heaton, item discrimination most generally "indicates the extent to which the item discriminates between the testees, separating the more able testees from the less able" (1989: 179).[63] In classical item analysis, the discrimination index of an item always includes a relation of scores on the item in question and the total scores of the overall testing population

---

[63]For a more detailed account of the concept of item discrimination c.f. e.g. McNamara (1996: 151).

(c.f. Alderson et al. 1995: 81f.). Although discussing the exact discrimination indices for the items analysed here would exceed the scope of this paper, it is nevertheless fairly obvious that [5.5.1.2.VII] does not have a high potential to discriminate between those subjects who possess greater and those with lower competencies in the use of prepositions. This can be said to be first of all due to the fact that the item provides test takers with a proportion of the word to be completed that can be hypothesised to almost automatically evoke the preposition *during* as the correct solution. The problem is further increased by the fact that in English the number of prepositions other than *during* and words in general that start with the letters *dur* is very limited or, in the case of prepositions, zero. It can therefore be assumed that the C-test format can be a useful tool for the testing of competence of English prepositions but that this usefulness depends to a large extent on the respective prepositions that are under focus.

For being able to further investigate the usefulness of the C-test as a means to test grammatical competence a closer look at section 2.3 of the TUC-PT, a text-based gap filling task which employs a multiple choice format will be taken. This will serve as a basis for comparing what has been discussed so far with reference to the grammar C-test in the WBPT with a technique that is more traditionally used for testing of grammar topics such as prepositions and tenses.

5.5.2 C-Testing versus Text-Based Multiple Choice Gap Filling: Comparison of Findings with Section 2.3 of the TUC-PT

As mentioned before, one of the advantages of the C-test format is the fact that test takers' decisions for a solution is not as restricted to receptive skills as in many other formats typically used for the testing of grammar, such as the multiple choice method (c.f. Heaton 1988: 42f.). Nevertheless, the above analysis showed that the C-test can at instances be problematic especially when certain areas of language proficiency, such as competence in the area of grammar, and not language proficiency in general are to be tested. It can be the case that for some C-test items a definite classification of the grammar point(s) that are being tested is impossible. This is certainly not the case with multiple choice gap filling tasks where a clear focus on grammar can be achieved via the selection of items. Furthermore, via the selection of distractors, the format provides the possibility of testing contrasts between particular grammatical phenomena such as tenses or prepositions. This is probably one of the reasons why multiple choice test tasks are one of the "most common types of objective items used to test awareness of the grammatical features of" languages (Heaton 1988: 34).[64]

---

[64]This statement clearly points to the restriction of the format to receptive as opposed to productive skills.

In the following, a closer look will be taken at section 2.3 of the TUC-PT, a text-based multiple choice gap filling task. As far as the format of this task is concerned, it can be said that section 2.3 contains a number of similarities to section one of the WBPT. Like the latter, the former includes 25 gaps with three alternatives the testees have to choose the correct solution from. This section focuses on a variety of grammar topics, which include participle constructions, relative clauses, conjunctions and linking expressions etc., the following analysis will however be restricted to the testing of tenses and prepositions in section 2.3 of the TUC-PT. In contrast to sections 2.1 and 2.2 of the TUC-PT, section 2.3 presents items within a longer coherent text, i.e. although the text that is the basis for section 2.3 does not seem to contain a very high degree of authenticity, it nevertheless uses materials that present certain grammar problems within an amount of context that reflects real language use to a larger extent than the single and more or less isolated sentences that form the basis for sections 2.1 and 2.2. However, as aspects of the influence of the context on the testing of tenses and prepositions have already been discussed above, the following parts of this paper will not be primarily concerned with the influence of contextual features.

5.5.2.1 Testing Tenses

As has been mentioned within the analysis of the potential of the C-test for the testing of tenses, one of the most obvious disadvantages of the format is the fact that for many items it is not clear whether they test competence in the use of tenses or simply the knowledge of rules for their formation (c.f. e.g. items [5.5.1.1.I] and [5.5.1.1.IV]). For the testing of tenses in section 2.3 of the WBPT, this does not seem to be a problem that is of primary importance. In total section 2.3 contains nine items that can be described to explicitly focus on the testing of tenses. Obviously, for the construction of a multiple choice test task, it is not necessary to leave parts or halves of the constructions under focus standing. Therefore, the format does not face testers with the problem of items calling too obviously for the choice of a particular letter or combination of letter which might in turn lead to the formation of a correct tense form no matter if the test takers are actually familiar with the rules for its use.

For all of the items of section 2.3 of the TUC-PT that primarily focus on the testing of tense-related competencies it can be said that the respective grammar point or tense that is under focus is clearly determinable. For item one of section 2.3 this focus lies upon the use of the simple past as opposed to other verb tenses or aspects expressing a temporal relation to the past, namely the past progressive and the present perfect.

> [5.5.2.1.I]
>
> Some time ago [1] I *was spending / **<u>spent</u>** / **have spent*** three weeks in Barcelona.

[5.5.2.1.I] was answered incorrectly by 34% of the total testing population. 11% decided for alternative one, the past progressive form and 23% of the testees chose alternative three the present perfect form of the verb *to spent*. This implies that the contrast between the simple past and the present perfect seemed to be more difficult to recognize for the subjects that were analysed that differences in the use of the simple past as opposed to the past progressive. The answer key for section 2.3 of the TUC-PT explicitly specifies alternative two, the simple past form, as the only correct answer to [5.5.2.1.I]. This can be justified on the basis of a definition of the simple past, or past tense, provided by Quirk et al. According to them, the tense "most commonly [...] combines" the following "two features of meaning"

> (a) The event/state must have taken place in the past, with a gap between its completion and the present moment.
> (b) The speaker or writer must have in mind a definite time at which the event/state took place. (1985: 183).

The most explicit way of emphasizing the second feature is a combination of a past tense verb form with "past time-position adverbials such as *last week*, *in 1932*, *several weeks ago*, *yesterday*, etc" (ibid.). Although "it is not necessary [...] for the past to be accompanied by an overt indicator of time" (184), the incorporation of the expression *some time ago* within the immediate context of [5.5.2.1.I] seems to serve the primary purpose of making the item unambiguous as it functions as an indicator of a "DEFINITE PAST" (ibid.) reference which is one of the most common uses of the simple past. In spite of the fact that a definite identification of the sources of the text and the sentences used as a basis for the grammar sections of the TUC-PT is not possible, it seems reasonable to suggest that they are not, or only to a fairly restricted extent, authentic. This shall be exemplified in closer detail in the course of this section. Therefore, it can be hypothesized that the incorporation of the time reference *some time ago* was consciously included in order to avoid ambiguities.

As has been mentioned within the discussion of the C-test format, the provision of a certain part of the structures to be reconstructed by the test taker might restrict the possibility for alternative solutions to an item. Generally speaking, this can be viewed as an advantage of the format as it normally leads to answer keys that can be established relatively easily which in turn renders the scoring procedure very objective and

economic and makes the technique especially useful for computer-based testing contexts. Besides this advantage, it needs to be maintained that the provision of parts of a tense form might make it very difficult to focus on contrasts between different tense and/or aspect forms, their respective meanings and uses. Such a clear focus on contrasts is however possible and at the same time a rather natural characteristic of the multiple choice method. In all cases, the provision of distractors obviously implies a certain focus on differences between the respective grammatical forms that are presented and their uses. With regard to the influence of the distractors on test objectivity this implies that the distractors need to be very carefully chosen a difficulty that is not implied in the process of developing C-test items.

One of the contrasts focused on by [5.5.2.1.I] is that between the simple past and the past progressive. As mentioned above, the expression *some time ago* most clearly calls for the use of simple past tense. The use of this tense implies a strong reference to the action of spending some time in Barcelona "as a whole" (Quirk et al. 1985: 197) and emphasises the temporal "gap between its completion and the present moment" (183). In contrast to that, the use of a past progressive verb form would put a main emphasis on the "activity" of spending time in Barcelona "in progress" (197). Especially as the whole text used in section 2.3 gives a detailed description of the period of the writer's staying in Barcelona, the places that were visited etc. it could become possible to hypothesize that the use of a past progressive form of the verb could under certain circumstances also have been possible in the context of [5.5.2.1.I]. However, the incorporation of the time adverbial clearly forbids this possibility and serves as a very strong cue for finding the correct solution to the item.

A second contrast, [5.5.2.1.I] puts special emphasis on is one of the "most problematic" ones related to tense and aspect, namely that "between simple past and present perfective" (Quirk et al. 1985: 189). Generally speaking, there exist a number of constructions in which "either the present perfective or the simple past can appropriately used", although the use of the two leads to different shades of meaning, i.e. "the present perfective relates the action more directly to the present time" (192). However, as soon as a "time-position adverbial is added [...], then the finite verb phrase has to be in the past tense [...] rather than the perfective" (191). On the ground of this, it becomes possible to note that the incorporation of *some time ago*, has a very crucial role to play in the construction of the item as it serves as the primary means of making contrasts between the uses of the three tense/aspect forms provided as alternatives explicit. Therefore, the item can be said to very clearly focus on fundamental contrasts with reference to tense and aspect in English. Especially the possibility of using the

simple past and the present perfect within one and the same construction that has been referred to by Quirk et al. (c.f. 1985: 192), hints to the fact that test items focusing on the use of different tenses/aspects can under certain circumstances allow more than one possible solution. The fact that the grammar sections of the TUC-PT are objective test components which generally do not allow for alternative solutions to the items, seems to make it necessary to construct items in such a way that alternative solutions become more or less impossible. This is achieved in [5.5.2.1.I] via the incorporation of a definite time reference. In comparison to the C-Test in the WBPT, it becomes possible to state that in this section such means of avoiding ambiguity were not consciously included. Rather, the format tries to avoid ambiguities by providing a more or less high, at times too high, proportion of the tense form to be supplied.

Although the assumption that section 2.3 of the TUC-PT includes a number of cues that were incorporated to serve the purpose of making items unambiguous cannot doubtlessly be proven, it can be further underpinned if one takes a look at item twelve of this section.

[5.5.2.1.II]

A few days later I [12] *was receiving / had received / <u>received</u>* a call from some acquaintances

---

Like [5.5.2.1.I], the immediate context of item contains a definite time reference which clearly calls for the selection of the simple past form *received* as the only possible solution. Compared to [5.5.2.1.I], the error rates for [5.5.2.1.II] are considerably lower. The item was answered incorrectly by 11% of the overall target group, with 6% and 5% deciding for alternatives one and two, respectively. Like [5.5.2.1.I], the item focuses on contrasts between the simple past and the past progressive. Furthermore, the past perfect is presented as a further alternative to the correct solution. The identification of the past progressive form as incorrect in the context of [5.5.2.1.II] can be justified on the basis of exactly the same facts that were cited with reference to [5.5.2.1.I], i.e. the definite time reference *a few days later* makes the use of the progressive virtually impossible. The past perfect form can also be identified as incorrect as the context provided by the whole text that is the basis for section 2.3 as well as the sentence and the paragraph the item is part of do not contain information that would justify the use of a past perfect verb form here. In other words, the past perfect could only be used if the context of the item contained "a time of orientation in the past", the action expressed by the verb phrase could be described as "anterior to" (Quirk et al. 1985: 196).

As has been mentioned with regard to [5.5.2.1.I], the inclusion of a definite time reference into the context of the item serves as the primary cue for making the decision between several alternative tense forms unambiguous. From the fact that [5.5.2.1.II] is already the second example for which this is the case, it can be hypothesized that the time expressions in question were intentionally included by the test developer(s) into the contexts of [5.5.2.1.I] and [5.5.2.1.II]. Although this seems a very useful technique for making section 2.3 more objective, it can be claimed that it makes the text that is the basis for the task appear rather inauthentic. It can be assumed that both, [5.5.2.1.I] and [5.5.2.1.II] very explicitly focus on contrasts in the use of tenses/aspects that express particular past time references. The items obviously make use of the potential of the multiple choice method to test competence with regard to contrasts between different grammatical forms and their uses– a possibility which is not to the same extent provided by the C-test. Furthermore, the degree of authenticity seems to be reduced in favour of an increase of test objectivity which is another aspect that can not to the same extent be found in the C-test as this format tries to incorporate authentic language materials to as high a degree as possible.

As could for instance be seen from the analysis of section one of the WBPT, the use of a high amount of authentic coherent materials might imply that the range of grammatical features that can be tested by a respective task might be restricted (c.f. chapter 4.3.1.1). In section 2.3 of the TUC-PT the range of different tense and aspect forms is more extended, i.e. some of the items focus on the use of constructions which seem relatively difficult to test on the basis of more authentic materials. One example for this is item 23 of section 2.3.

[5.5.2.1.III]

Now I wonder what changes [23] **have occurred / <u>will have occurred</u> / will occur** by the next time I have a chance to go to Barcelona

---

This item was answered incorrectly by a very high proportion of the target group; 45% of all subjects decided for one of the distractors, with a high majority of 40% choosing alternative three and only 5% selecting the present perfect form *have occurred*. The answer key for section 2.3 identifies alternative two, a construction that is often referred to as the future perfect form of the verb[65], as the only possible solution. From a formal

---

[65]A number of researchers do not distinguish an independent future tense in English (c.f. e.g. Huddleston 1984: 133 or Huddleston & Pullum 2002: 208ff.) and the label *future perfect* is only used by a small amount of grammar reference works (e.g. Greenbaum 1996). For reasons of conciseness the construction

point of view the correct solution to [5.5.2.1.III] can be described as a combination of the modal verb *will* with a full verb that has a (present) perfective aspect. Corpus-based analyses have shown that modals such as *will*, which can imply "future time connotations [...] very rarely occur in the perfect aspect" (Biber et al. 1999: 297ff.). Nevertheless, [5.5.2.1.III] focuses on such a construction that is rather infrequent in real-life language use. This can probably be explained by the fact that the TUC-PT is a placement instrument for students of university subjects that put a special emphasis on a detailed knowledge of the grammatical structures of the English language in its entirety. It can therefore be claimed that the purpose and the target group of the test are some of the possible reasons why section 2.3 uses materials that seem to contain a fairly low amount of authenticity.

The fact that the future perfect, or the combination of *will* with perfective aspect, is a rather rare phenomenon in English is also to a certain extent reflected in many grammar reference works; the construction is mostly not explained in closer detail and there are few grammarians who make an explicit reference to it. One of them is Greenbaum who describes the future perfect as "referring to the past within the future" (1999: 274). This reference to a past event or state within the future is also contained in [5.5.2.1.III]. Here the correct solution, the future perfect form, is to denote the fact that the occurrence of particular changes will be completed by a particular time in the future; a reference to this time in the future is explicitly made with the help of the expression *by the next time I have a chance to go to Barcelona*. Furthermore, the use of the preposition *by* implies a reference to the end point of a time period and thus implies certain aspects of completion. This is also reflected in the following definition of the use of the future perfect provided by Swan. "We can use the future perfect to say that something will have been done, completed or achieved <u>by</u> a certain time in the future" (1998: 224, my underlining).

This aspect of completion that is characteristic for the future perfect would not be expressed by the use of a combination of *will* and infinitive, i.e. alternative three. Quirk et al. describe such constructions "as the closest approximations to a colourless, neutral future", which nevertheless frequently include semantic aspects of prediction and occasionally also volition when referring to the future (1985: 213; 228f.). The aspect of "'past in the future'" (191), which is typically expressed by the future perfect, is however not implied in such VPs. Therefore, alternative three can doubtlessly be justified as an incorrect alternative to [5.5.2.1.III]. It can furthermore be claimed that it is again an explicit indication of time, i.e. a combination of *by* with an expression

---

under focus here will nevertheless be treated under this name in the course of this section.

referring to a certain point in time in the future that functions as a major cue for the identification of the only possible correct alternative.

In contrast to alternative three which lacks a reference to the past, alternative one can be described as incorrect as it does not include the aspects of futurity that need to be denoted by the VP. The present perfect does not include future meanings in English. In contrast to that, however, the German *Perfekt*, mostly in combination with a further expression of future time, can be used to denote events or situations that are expected to be completed at some point in the future (Sommerfeldt & Starke 1998: 69). "Das Perfekt stellt ein erwartetes, also zukünftiges Geschehen dar. Zur Realisierung der Bedeutung 'zukünftig' sind weitere temporale Mittel notwendig. In dieser Bedeutung ersetzt das Perfekt in zunehmendem Maße das Futur II" (ibid.). One of the possible meanings of the *Futur II* is the fact that it can express "die Vermutung über ein in der Zukunft als abgeschlossen gedachtes Geschehen" (67). In this respect as well as with regard to its formal characteristics, the *Futur II* in German can be described as the closest equivalent to the English future perfect. The fact, however, that the *Futur II* is nowadays rather infrequently used in this sense in German (70) and that it is commonly replaced by the *Perfekt* might be one of the reasons why alternative one was attractive to at least a small number of testees. It can thus be claimed that by presenting alternative one as a distractor for [5.5.2.1.III] a certain emphasis is put upon potential problems that might originate from an interference of the German mother tongue of most of the testees. However, though very probable, it cannot be proven without further investigation whether this was the primary reason for including distractor one into the list of alternatives for [5.5.2.1.III]. What can more definitely be claimed is the fact that the item requires testees to recognize a combination of aspects of completion and futurity and that these aspects are made very explicit with the help of a particular expression of time that is included into the context of the item. This again underscores the assumption that the role that authenticity of materials has to play in section 2.3 of the TUC-PT seems to be fairly low when compared to section two of the WBPT. Therefore, it can be claimed that the text-based multiple choice task seems to include a high amount of conscious adaptation or even construction of input materials in order to make the items unambiguous whereas for the C-test adaptation of materials was kept as low as possible. It is very striking that the immediate context of a very high amount of the tense items in section 2.3 of the TUC-PT contains an explicit reference to time which makes the choice of the respective correct alternative for all of them very clear and unambiguous. Below, a further example, which shall only be touched briefly, is item two of section 2.3

125

> [5.5.2.1.IV]
>
> Two years had passed since the Olympic Games [2] *were taking place / had been taking place / <u>had taken place</u>* and the city had changed greatly.

The construction contains two past perfect forms, in addition to the one being the correct solution to the item, especially the form *had changed* can be judged to express a time relation that is more or less parallel to that which needs to be expressed by the correct solution to item two. This can most clearly be seen by the conjunction *and* which joins the two VPs. Furthermore, the expression *two years had passed since...* makes a definite reference to a time earlier than the visit of the writer to Barcelona. As this visit lies in the past, the use of the past perfect in [5.5.2.1.IV] is very explicitly called for by the context of the item.

As could be seen from the above discussion, time references whether they are expressed with the help of time adverbials, conjunctions or the tenses of verb forms within the context of particular items play a very important role for items in section 2.3 of the TUC-PT that focus on tenses. Naturally, this can not be assumed to be the case for items that deal with prepositions. In the following, the only preposition item that is contained in section 2.3 will be discussed with reference to the way the item is designed and constructed and the contrasts that it focuses on.

5.5.2.2 Testing Prepositions

Section 2.3 of the TUC-PT contains only one item, i.e. item eleven, which can be identified to explicitly focus on prepositions.

> [5.5.2.2.I]
>
> so we returned to a place we knew [11] *<u>close to</u> / in the near of / near of* the Old Quarter.

[5.5.2.2.I] yielded an error rate of 23% with 17% and 6% of the testing population deciding for alternatives two and three, respectively.

It has been observed above that for the testing of tenses, section 2.3 puts a special emphasis on the subjects' competence with regard to contrasts between the use of different tense forms and to the design of items that are unambiguous. As has been mentioned before, the C-test format does not imply as many possibilities to focus on contrasts between different forms. In the following the contrasts with regard to the use of prepositions that are focused on by [5.5.2.2.I] will be discussed in closer detail. The item puts a main emphasis on the use of prepositions that refer to space or location. Both the correct alternative, *close to*, as well as the preposition *near*, which is part of distractor three, *near of*, are most commonly used with such meanings (Quirk et al. 1985: 679). Like *close to*, *near (to)* "has locative meaning and the phrases it heads are [...] like uncontroversial locative PPs such as *in the pool* or *beyond the city*" (Huddleston & Pullum 2002: 609). The same could be claimed for the NP *the Old Quarter* in [5.5.2.2.I]. Nevertheless, alternative three can doubtlessly be identified as incorrect in the context of [5.5.2.2.I] as *near* can only be accompanied by *to* as a second preposition forming the head of the following NP (ibid.). Moreover, the construction *near of* is not normally used in English, or can even be said not to exist in the language at all.

The word *near* can be a member of a variety of word classes, it can be a verb, an adverb, a preposition, or an adjective (c.f. *OED*: s.v.) it can, however, not be used as a noun. It is for this reason that alternative three, *in the near of*, can also be identified as being incorrect as an *of*-construction generally includes a noun. Therefore, an expression like this would hardly ever be used in English. In contrast to English, genitive constructions like *in der Nähe des (Old Quarter)* are relatively common in German which hints to the fact that [5.5.2.2.I] is a further example of items in section 2.3 of the TUC-PT that focus on potential L1 interferences. The fact that two was chosen by the majority of test takers who answered [5.5.2.2.I] incorrectly implies that a certain number of them might have used structures that would be appropriate in their German mother tongue as a basis for their choice of alternatives to [5.5.2.2.I].

It can be claimed that although the two distractors to [5.5.2.2.I] are hardly ever used in English they both include the most important semantic information that is called for by the item. As both, alternatives two and three contain the word *near*, which implies meaning associations that are very similar to those of the correct solution *close*

to (Lindstromberg 1997: 141) the item cannot be described to focus primarily on semantic contrasts between different English prepositions. It rather seems to put particular emphasis on test takers' competence with regard to the composition of (complex) prepositions and to questions of certain constructions being used in the language at all. As has been claimed with regard to C-testing, this format implies a certain potential for testing competence with regard to the meaning, composition and use of a variety of prepositions. It does so, however, only in cases where the amount and nature of letters provided are suitable for this purpose. Furthermore, there is only a limited degree of control over the contrasts that are focused on by C-test items possible. For [5.5.2.2.I], however, as well as any other multiple choice item, it can be claimed that there is a clear identification of the contrasts that are tested possible. It furthermore seems that the item is constructed in such a way that both distractors can be described as genuinely incorrect. In contrast to most other preposition items that have been discussed above, the distractors to [5.5.2.2.I] cannot primarily be described to be incorrect on the grounds of their semantics but by the fact that they are hardly existent in the English language.

As [5.5.2.2.I] is the only preposition item in section 2.3 the above assumptions do not allow any generalizations concerning the testing of prepositions in this part of the test. Nevertheless, a comparison of [5.5.2.2.I] with the preposition items in sections 2.1 and 2.2 of the same test implies that unambiguous test items can be achieved via a variety of different techniques. Whether the one used for [5.5.2.2.I] is very useful with regard to aspects of real-life communication remains, however, questionable as the item calls for the testees' dealing with structures that they will probably not or very rarely have to deal with when using the English language. Comparing this to the C-test, it can be claimed that for items such as [5.5.1.2.VI] which do not call to obviously for a certain type of preposition, it can at instances be relatively difficult for answer keys to account for different possible solutions. Besides not allowing for a clear focus between different types of prepositions, C-tests also need to use different techniques for rendering items unambiguous, such as the provision of a higher number of letters or – what would in most cases be more useful – the careful design and revision of answer keys so that every possible alternative can be accounted for.

5.5.2.3 Summarising the Comparison

The comparison of section two of the WBPT with section 2.3 of the TUC-PT proved to some extent Coleman's claim that for C-tests a clear identification of the particular language and/or grammar point to be tested can be problematic (c.f. 1994:

218). Especially with regard to the testing of tenses questions whether a particular C-test item focuses on competence in the use of a tense form or simple knowledge of the principles of its formation. As English verb tenses constructions often consist of longer, multi-word constructions which makes it possible to claim that the deletion of half of a word within such a construction leaves the testee with too much of the information necessary to reconstruct the respective form. competence in use of tenses or simply knowledge of rules of formation of tenses – especially problematic for tenses, as these (unlike prepositions) most often consist of longer constructions, i.e. deletion of half of a word could leave the testee with too much information (c.f. e.g. [5.5.1.1.I]). In section two of the WBPT, a specific focus upon the use of tenses within different contexts only seems possible for tense forms for which the deletion of parts of a word leaves a stem that provides enough space for different tense forms that actually exist in the language (c.f. e.g. [5.5.1.1.II]). Some of the items in this section can be said to primarily focus on syntactic requirements, such as the need for a finite verb within a clause, rather than the use of tense forms (c.f. e.g. [5.5.1.1.IV]).

The identification of the grammatical problem being tested by the C-tests in section two of the WBPT seem slightly less problematic with regard to the testing of prepositions, at least with regard to a certain group of simple prepositions (c.f. e.g. [5.5.1.2.III]). Simple prepositions for which it is possible to provide test takers with a relatively small amount of letters, seem to have the potential to focus on a variety of semantic and/or syntactic properties of the word class. As could be seen with regard to [5.5.1.2.VI], this can also lead to problems regarding the appropriateness of more than one answer which might in turn have to lead to a revision of answer keys. As was the case for the testing of tenses, it can be claimed that C-test items in section two of the WBPT which focus on prepositions that do not share some of the formal characteristics of a high number of members of the word class provide testees with too much information (c.f. [5.5.1.2.VII]). Such items can be described as not focusing on competence with regard to the use of prepositions but simply on knowledge of the lexical forms.

Apart from the fact that an explicit focus on a particular grammar point within the C-test format can at instances be problematic, the technique does not directly test subjects' understanding of contrasts between different tense forms and prepositions. As has been mentioned earlier on, such a focus on the use of different forms within one item is however one of the most important characteristics of the multiple choice method. With regard to section 2.3 of the TUC-PT it can be assumed that, in contrast to section two of the WBPT, the task allows as clear identification of the grammatical

topic under examination as well as an emphasis on different tense and preposition constructions. It has nevertheless to be claimed that, although the format employed in section 2.3 seems to be more applicable to the explicit testing of grammatical competence, the task as such could be considerably improved if items were constructed in a more natural way or with the help of more authentic base materials.

6. Conclusion

In this paper, a detailed formal and linguistic comparison of a computer-based and a paper-and-pencil-based placement test with special reference to the explicit testing of grammar, precisely tenses and prepositions was provided. The formal characteristics of both tests were investigated on the basis of a test classification framework provided by Bachman (1990) which showed that the five features contained in this categorisation provide a particularly useful tool for the description of language tests along different but complementary dimensions. Furthermore, it was found out that although the TUC-PT and the WBPT share a number of important characteristics, especially with regard to their content, frame of reference and scoring procedure, differences with regard to test purpose and the test methods being used were detectable.

The linguistic comparison focused on two overall criteria, first, the influence of the context of coherent texts as opposed to single sentences on multiple choice testing of tenses and prepositions, and second the applicability of the C-test method to the assessment of these two areas of English grammar when compared to a text-based multiple choice task.

With regard to the first criterion, it could be found that for the items in sections 2.1 and 2.2 of the TUC-PT analysed by this paper, the context of the sentence or a small set of single sentences was provided and constructed in such a way that the items were not rendered unambiguous. Likewise, the description of items in section one of the WBPT showed that the most important information necessary for deciding on a particular alternative, was in most cases provided by their immediate context. Nevertheless, the comparison of section one of the WBPT with sections 2.1 and 2.2 of the TUC-PT showed that a broader context, ideally in the form of an authentic, coherent text, has an important supportive function as it implies crucial information which can be used by the test taker in order to identify a correct or most appropriate alternative. This became particularly notable with regard to choices, such as the use of tenses in reported speech constructions that are not only dependent upon grammatical but also pragmatic aspects. As the testing of pragmatic competencies is generally not conducted with the help of testing techniques such as the ones analysed in this paper, it needs to be

underscored that the provision of a rather high amount of context within explicit measures of grammatical competence is probably less crucial for such testing situations than it is for more integrative and/or task-based tests. Nevertheless, Heaton's assumption that the presentation of more or less decontextualized test items can create an impression of language that does not correspond to its actual nature as a means of conveying information of various kinds within a variety of realistic and authentic situations (cf. 1988: 36) could be proven. Comparing items from sections 2.1 and 2.2 of the TUC-PT with those from section one of the WBPT, it became notable that the former appeared at instances rather inauthentic or even constructed. Nevertheless, as the analysis of section one of the WBPT showed, authentic language contexts, such as news articles, can also have the disadvantage of imposing certain restrictions on the grammatical contents that can be tested on the basis of such materials (c.f. Heaton: 1988: 36). One of the crucial findings of this paper is, thus, the assumption that a broad, coherent, possibly authentic language context can be a very useful means for bringing a test method such as multiple choice which is characterised by a number of restrictions and a certain degree of artificiality closer to real-life communicational contexts. However, the materials used as a basis for multiple choice test tasks need to be chosen with particular reference to the grammatical contents that are to be tested and treated with special care.

With regard to the second criterion, the usefulness of the C-test for the explicit testing of grammar in comparison to the multiple choice technique, it could be found that the C-test, even in a modified version that puts a main focus on grammatical contents, can lead to a number of problems. In this respect, two main aspects could be identified. First, the C-test does at many instances make an identification of the particular grammar point to be tested difficult or even impossible. Second, an explicit focus on contrasts between the use of different grammatical forms belonging to the categories of tenses or prepositions within one and the same item can hardly be achieved. Rather the tester sometimes does not seem to have a high amount of control over the respective solutions a C-test item evokes. In contrast to that, the multiple choice method allows the tester to define the grammatical contents under focus in a very straightforward and clear-cut way. The comparison of section two of the WBPT with section 2.3 of the TUC-PT showed that for the latter, the respective contents and contrasts to be assessed were very clearly definable. I needs to be underlined that the C-test is generally described as aiming at an integration of several skills within one and the same test task, or a measure of global language proficiency (Grotjahn 1995: 37). It can be said to focus on a combination of different skills, such as reading, as well as writing

and the ability to make use of natural redundancies within language. Via a number of modifications of the general construction principles for C-tests, an attempt was taken to adapt this technique, which obviously focuses on a less restricted range of abilities than the multiple choice method, to the area of explicit grammar testing. The problems identified above, however, seem to imply that when compared to the multiple choice technique, the C-test is probably less useful as a means to explicitly test grammatical competencies with regard to tenses and prepositions, at least in contexts such as the testing of prospective students of English and American Studies or a related subject as this target group generally needs to possess very detailed knowledge of the grammatical structures of the language. Although the multiple choice technique seems to have a number of advantages when compared to what has just been mentioned, it can be claimed that especially with regard to the testing of tenses, section 2.3 of the TUC-PT includes an amount of cues used to render items unambiguous which makes the task or parts of it appear relatively constructed and inauthentic. With regard to the only preposition item in section 2.3, it seems that ambiguities were first of all avoided via the provision of alternatives. However, the fact that the task contains such a low number of items focusing on English prepositions does not allow any generalisations concerning this topic. At least with regard to the testing of tenses in section 2.3, it can be claimed that, although the multiple choice technique can be described as a useful means of explicitly testing grammatical competence, it is not only the choice of a particular test method which needs to be carefully considered when designing grammar test tasks, or language tests in general, but also the way materials are chosen and items are developed.

From this it can be inferred that the design of items is a very crucial aspect related to the analysis and comparison of language tests and it is for this reason that the present paper investigated the items discussed in a very detailed way focusing on the grammatical contents and rules being tested, and ,where applicable, a discussion of the respective alternatives as well as individual problems on the side of the test taker. Although, especially because of the selection of base materials, the items in the WBPT that were discussed here seem to focus on more realistic grammar problems than some of the items in the grammar component of the TUC-PT it needs to be clearly emphasised that both formats focus on the explicit testing of grammar which has lost much of its influence in the course of a growing interest in the principles of communicative language testing (c.f. Rea-Dickins 2001: 23; 29). Taking into account the target group of both tests and the special requirements that can be identified with regard to their grammatical competencies it becomes possible to claim that the explicit

testing of grammar is in any case a necessary component of placement tests for prospective students of university subjects that are strongly related to the English language and particularly the ability to at the same time use and analyse this language.

However, grammatical competence, needs to be treated as only one component of language proficiency or communicative language ability as defined by Bachman (cf 1990: 81-190). This is also reflected by the fact that the TUC-PT contains components that do not focus on the explicit testing of grammar but on the assessment of a combination of reading, writing, listening, and speaking skills in combination with grammatical, as well as other competencies. The WBPT clearly needs to be understood as a practise version for the sections of this relatively broad test battery which focus on grammatical (and textual) competencies as just one component of communicative language ability.

This restriction is, amongst other aspects, due to the fact that especially for tests of grammar, computer-based formats are considered rather useful. The grammar sections of the TUC-PT which are all objectively measurable, therefore offered a very interesting basis for the design of a web-based practise placement test component. Being a web-based test instrument, the WBPT can easily be accessed by anybody who wishes to practise for the TUC-PT or to improve his/her English grammar competence in general. The number of subjects who have taken the test so far is very encouraging and from this it follows that the test bears a high potential for future applications either as a practice test or a computer-based substitute for the grammar sections of the TUC-PT. Apart from focusing on the research questions addressed above, the analysis provided in the present paper furthermore yielded some interesting insights with regard to possible future improvements of the WBPT, such as the restriction of alternative versions in each section in order to achieve higher equivalence among them, or a further analysis of the C-test section with reference to possible alternative solutions which need to be included into the answer key. Moreover, test taker data for all the subjects who have taken and take the WBPT are stored in a database and can therefore easily be scored and accessed. The WBPT, adaptations, or sections of it, could thus serve as a useful basis for further research into the testing of English grammar and accordingly constitute what Bachman describes as one very important use of language tests, namely an indicator "of abilities or attributes that are of interest in research on language, language acquisition, and language teaching" (1990: 54).

7. References

7.1 Primary Sources

Ancient Athens Caught in New Light for Olympics (Ancient Athens I). July 19, 2004.
Retrieved 21 July, 2004 from
http://sportsillustrated.cnn.com/2004/olympics/2004/07/19/bc.eu.spt.oly.athens.p
arthenon.ap/index.html?cnn=yes.

Ancient Indians Made 'Rock Music' (Ancient Indians I). March 19, 2004. Retrieved
August 25, 2004 from
http://news.bbc.co.uk/go/pr/fr/-/1/hi/sci/tech/3520384.stm.

Brenda Fassie: A Very Human Hero (Brenda Fassie I). October 5, 2004.
Retrieved October 24, 2004 from http://news.bbc.co.uk/go/pr/fr/-
/2/hi/africa/3700309.stm.

Country Profile: United Kingdom (Country Profile United Kingdom I).
February 29, 2004. Retrieved August 28, 2004 from
http://news.bbc.co.uk/go/pr/fr/-
/2/hi/europe/country_profiles/1038758.stm.

Early UK Springs Become 'Normal' (Early UK Springs I). March 12, 2004.
Retrieved August 25, 2004 from http://news.bbc.co.uk/go/pr/fr/-
/2/hi/science/nature/3502772.stm.

Jane Austen – The Hottest Writer in Hollywood (Jane Austen I). In Soars, & Soars
(1998: 82-83).

Snapshots of Britain: A Sense of Identity (A British National Identity I). In McDowall
(1999: 10-12).

Teacher Standards Held for Some Schools (New Teacher Standards I). July 16, 2004.
Retrieved August 25, 2004 from
http://www.cnn.com/2004/EDUCATION/07/16/rural.school.ap/index.html

7.2 Textbooks

McDowall, D. (1999). *Britain in Close-Up: An In-Depth Study of Contemporary Britain*
(new ed.). Harlow: Longman.

Soars, J., & Soars, L. (1998). *New headway upper-intermediate English course*. Oxford:
Oxford University Press.

7.3 Secondary Sources

Alderson, J.C. (1996). Do Corpora Have a Role in Language Assessment?. In J. Thomas & M. Short (Eds.), *Using corpora for language research* (pp. 248-259). London: Longman.

Alderson, J. C, Clapham, C, & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Alderson, J. C., & Banerjee, J. (2001). Language Testing and Assessment (part I). *Language Teaching*, 34, 213-236.

Alderson, J. C., & Banerjee, J. (2002). Language Testing and Assessment (part II). *Language Teaching*, 35, 79-113.

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bachman, L. F. (2000). Modern Language Testing at the Turn of the Century: Assuring That What We Count Counts. *Language Testing*, 17(1), 1-42.

Bachman, L. F., & Palmer A. S. (1996). *Language Testing in Practice*. Oxford: Oxford University Press.

Baker, D. (1989). *Language Testing: A Critical Survey and Practical Guide*. London: Edward Arnold.

Bugbee, A. C. (1996). The Equivalence of Paper-and-Pencil and Computer-Based Testing. *Journal of research on computing in education*. 28(3), 282-299.

Canale, M., & Swain, M. (1980). Theoretical Bases of Communicative Approaches to Second Language Teaching and Testing. *Applied Linguistics*, 1(1), 1-47.

Chalhoub-Deville, M. (1997). Theoretical Models, Assessment Frameworks and Test Construction. *Language Testing*, 14 (1), 3-22.

Chalhoub-Deville, M., & Deville C. (1999). Computer-Adaptive Testing in Second Language Contexts. *Annual Review of Applied Linguistics*, 19, 273-99.

Chemnitz University of Technology English and American Studies Kommentiertes Vorlesungsverzeichnis Wintersemester 2004/05. Chemnitz University of Technology, Department of English and American Studies.

Coleman, J. A. (1994). Profiling the Advanced Language Learner: The C-Test in British Further and Higher Education. In Grotjahn, R. (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen*. (Vol. 2), (pp. 217-237). Bochum: Brockmeyer.

Davies, A. (1990). *Principles of Language Testing*. Oxford: Blackwell.

Ellis, R. (1997). *Second Language Acquisition*. Oxford: Oxford University Press.

Fulcher, G. (1997). An English Language Placement Test: Issues in Reliability and Validity. *Language Testing*, 14 (2), 113-139.

Fulcher, G. (1999). Computerizing a Language Placement Test. *ELT Journal*, 53 (4), 289-299.

Grotjahn, R. (1995). Der C-Test: State of the Art. *Zeitschrift für Fremsprachenforschung*, 6(2), 37-60.

Heaton, J. B. (1988). *Writing English Language Tests* (new ed.). London, New York: Longman.

Huhta, A. (1996). Validating an EFL C-Test for Students of English Philology. In Grotjahn, R. (Ed.), *Der C-Test: Theoretische Grundlagen und praktische Anwendungen* (Vol. 3), (pp. 197-234). Bochum: Brockmeyer.

Klein, W. (1994). *Time in Language*. London, New York: Routledge.

Kunnan, A. J. (1999). Recent Developments in Language Testing. *Annual Review of Applied Linguistics*, 19, 235-253.

Lado, R. (1961). *Language Testing*. London: Longman.

Lindstromberg, S. (1997). *English Prepositions Explained*. Amsterdam, Philadelphia: John Benjamins Publishing Company.

Mair, C. (1995). *Englisch für Anglisten: Eine Einführung in die Englische Sprache*. Tübingen: Stauffenburg-Verlag.

McNamara, T. (1996). *Measuring Second Language Performance*. London, New York: Longman.

Oller, J. W. (1976). Evidence of a General Language Proficiency Factor: An Expectancy Grammar. *Die Neuen Sprachen*, 76, 165-74.

Oller, J. W. (1979). *Language Tests at School: A Pragmatic Approach*. London: Longman.

Palmer, F. (1988). *The English Verb* (2nd ed). London: Longman.

Palmer, F. (1990). *Modality and the English Modals* (2nd ed.). London: Longman.

Piper, A. (1983). A Comparison of the Cloze and C-Test as Placement Items. *The British Journal of Language Teaching*, 21, 45-51.

Rea-Dickins, P. (1991). What Makes a Grammar Test Communicative?. In J. C. Alderson & B. North (Eds.), *Language Testing in the 1990s: The Communicative Legacy*. (pp. 112-131). London: Macmillan.

Rea-Dickins, P. (1997). The Testing of Grammar in a Second Language. In C. Clapham & D. Corson (Eds.), *Language Testing and Assessment*. (Vol. 7), (pp. 87-97). Dordrecht: Kluwer Academic Publishers.

Rea-Dickins, P. (2001). Fossilisation or Evolution: The Case of Grammar Testing. In C. Elder, A. Brown, E. Grove, K. Hill, N. Iwashita, T. Lumley, T. McNamara, & K. O'Loughlin (Eds.), *Experimenting with Uncertainty: Essays in Honour of Alan Davies* (pp. 22-32). Cambridge: Cambridge University Press.

Spolsky, B. (1995). *Measured Words*. Oxford: Oxford University Press.

Tyler, A., & Evans, V. (2004). *The Semantics of English Prepositions: Spatial Scenes, Embodied Meaning and Cognition* (Reprinted ed.). Cambridge: Cambridge University Press.

Wall, D., Clapham, C., & Alderson, J. C. (1994). Evaluating a Placement Test. *Language Testing*, 11(3), 321-344.

Weir, C. J. (1990). *Communicative Language Testing*. Hemel Hempstead: Prentice Hall International.

Weir, C. J. (1993). *Understanding and Developing Language Tests*. Hemel Hempstead: Prentice Hall International.

Zelinsky-Wibbelt, C. (1993). Introduction. In C. Zelinsky-Wibbelt (Ed.), *The Semantics of Prepositions: From Mental Processing to Natural Language Processing* (pp. 1-24). Berlin, New York: de Gruyter.

7.4 On-line Sources

ALTE framework and can-do. Association of Language Testers in Europe (ALTE). Retrieved February 8, 2005 from http://www.alte.org/can_do/framework/index.cfm.

Bhela, B. (1999). Native Language Interference in Learning a Second Language: Exploratory Case Studies of Native Language Interference with Target Language Use. *International Education Journal*, 1 (1), 22-31. Retrieved April 13, 2005 from http://iej.cjb.net.

British and American Studies – Veranstaltungsangebot Wintersemester 2004/2005. Chemnitz University of Technology, Department of British and American Studies. Retrieved March 22, 2005 from http://www.tu-chemnitz.de/phil/english/chairs/cultstud/pages/4_03_prev_cour_g.html.

Brown, J. D. (1997). Computers in Language Testing: Present Research and Future Directions. *Language Learning and Technology*. 1(1), 44-59. Retrieved October 24, 2004 from http://llt.msu.edu/vol1num1/brown/default.html.

Dakota State University ESL Placement Test. Dakota State University, Madison SD. Retrieved November 15, 2004 from http://www.act.org/esl/descript.html.

Dakota State University Placement Test: Sample Questions. Dakota State University,
Madison SD. Retrieved, November 15, 2004 from
http://www.act.org/compass/sample/index.html.

Dresden University of Technology – Entry Test in Winter Semester 2005/2006.
Technische Universität Dresden, Institut für Anglistik & Amerikanistik. Retrieved
July 3, 2005 from http://www.tu-dresden.de/suliaa/angl0.htm.

Dresden University of Technology - Placement Test (WS 2004). Technische Universität
Dresden, Institut für Anglistik & Amerikanistik. Retrieved November 11, 2004
from http://www.tu-dresden.de/suliaa/angl0.htm.

Joyce, P. (n.d.). *Prepositions and their Usage.* University of Exeter, German
Department. Retrieved May 23, 2005 from
www.ex.ac.uk/~pjoyce/gramlect/gramlect9.ppt.

Lehrplan Gymnasium Englisch Klassenstufen 5 bis 10, Jahrgangsstufen 11 bis 12.
Sächsisches Staatsministerium für Kultus. Retrieved March 8, 2005 from
http://www.englisch-hilfen.de/lehrer/lehrplan_englisch_gymnasium.pdf.

Miyata–Boddy, N., & Langham, C. S. (2000). *Communicative Language Testing - An
Attainable Goal?* The British Council, Tokyo. Retrieved October 22, 2004 from
http://www.kasei.ac.jp/library/kiyou/2000/5.LANGHAM.pdf.

Placement Testing at Shippensburg University. Shippensburg University of
Pennsylvania. Retrieved November 15, 2004 from
http://www.ship.edu/~testing/head2.html.

Power, T. (2003). *Language Testing & Methods of Assessment.* Retrieved November 2,
2004 from http://www.btinternet.com/~ted.power/esl0712.html.

Raatz, U., & Klein-Braley, C. (n.d.). Introduction to Language Testing and C-Tests.
Retrieved December 15, 2004 from Universität Duisburg-Essen, Angewandte
Linguistik http://www.uni-
duisburg.de/FB3/ANGLING/FORSCHUNG/HOWTODO.HTM.

Roever, C. (2001). Web-Based Language Testing. *Language Learning and Technology.*
5 (2), 84-94. Retrieved October 24, 2004 from
http://llt.msu.edu/vol5num2/roever/default.html.

Wharton, S. (n.d.). *Characterising your Language Test.* Retrieved November, 2, 2004
from http://www.philseflsupport.com/testing1.htm.

Zhang, L. (n.d.). *Language Differences and Communication: Contrastive Analysis
Revisited.* Retrieved October 22, 2004 from The University of Western Ontario,
Department of Education
http://www.uwo.ca/sogs/academic/WJGR/2005/WJGR2005_v12_p113_Zhang.ht

m.

7.5 Grammar Reference Works

Biber, D, Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman Grammar of Spoken and Written English*. London: Longman.

Greenbaum, S. (1996). *The Oxford English Grammar*. Oxford: Oxford University Press.

Huddleston, R. (1984). *Introduction to the Grammar of English*. Cambridge: Cambridge University Press.

Huddleston, R, & Pullum, G. (Eds.) (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.

Leech, G., & Svartvik, J. (1994). *A Communicative Grammar of English* (2nd ed.). London; New York: Longman.

Quirk, R., Greenbaum, S, Leech, G., & Svartvik, J. (1974). *A Grammar of Contemporary English* (4th impr.). London: Longman.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.

Sommerfeldt, K.-E., & Starke, G. (1998). *Einführung in die Grammatik der deutschen Gegenwartssprache* (3rd ed.). Tübingen: Niemeyer.

Swan, M. (1998). *Practical English Usage* (new ed.). Oxford: Oxford University Press.

7.6 Dictionaries and Encyclopaedias

Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara T. (1999). *Dictionary of Language Testing (Studies in Language Testing 7)*. Cambridge: Cambridge University Press.

Johnson, K., & Johnson, H. (1998). *Encyclopedic Dictionary of Applied Linguistics*. Oxford: Blackwell.

Schröder, J. (1987). *Lexikon deutscher Präpositionen*. Leipzig: VEB Verlag Enzyklopädie.

Longman Dictionary of Phrasal Verbs (LDPV) (14th impr.). (1990). London: Longman.

The Longman Dictionary of the English Language (LDEL) (2nd ed.). (1991). Harlow: Longman.

The New Oxford Dictionary of English (NODE). (1998). Oxford: Oxford University Press.

The Oxford English Dictionary (OED). (1991). Oxford: Oxford University Press.

8. Appendices

Appendix I: Bachman (1990) Figures 4.2 and 5.1

Bachman (1990: 87) Fig. 4.2 *Components of Language Competence*

Bachman (1990: 119) Fig. 5.1 *Categories of Test Method Facet*

1 FACETS OF THE TESTING ENVIRONMENT

Familiarity of the place and equipment
Personnel
Time of testing
Physical conditions

---

2 FACETS OF THE TEST RUBRIC

Test organization
   Salience of parts
   Sequence of parts
   Relative importance of parts

Time allocation

Instructions
   Language (native, target)
   Channel (aural, visual)
   Specification of procedures and tasks
   Explicitness of criteria for correctness

---

3 FACETS OF THE INPUT

Format
   Channel of presentation (aural, visual)
   Mode of presentation (receptive)
   Form of presentation (language, nonlanguage, both)
   Vehicle of presentation ('live', 'canned', both)
   Language of presentation (native, target, both)
   Identification of problem (specific, general)
   Degree of speededness

Nature of language
   Length
   Propositional content
      Vocabulary (frequency, specialization)
      Degree of contextualization (embedded/ reduced)
      Distribution of new information (compact/ diffuse)
      Type of information (concrete/abstract, positive/negative, factual/counter-factual)
      Topic
      Genre
   Organizational characteristics
      Grammar
      Cohesion
      Rhetorical organization
   Pragmatic characteristics
      Illocutionary force
      Sociolinguistic characteristics

---

4 FACETS OF THE EXPECTED RESPONSE

Format
   Channel (aural, visual)
   Mode (productive)
   Type of response (selected, constructed)
   Form of response (language, nonlanguage, both)
   Language of response (native, target, both)

Nature of language
   Length
   Propositional content
      Vocabulary (frequency, specialization)
      Degree of contextualization (embedded/ reduced)
      Distribution of new information (compact/ diffuse)
      Type of information (concrete/abstract, positive/negative, factual/counter-factual)
      Topic
      Genre
   Organizational characteristics
      Grammar
      Cohesion
      Rhetorical organization
   Pragmatic characteristics
      Illocutionary force
      Sociolinguistic characteristics

Restrictions on response
   Channel
   Format
   Organizational characteristics
   Propositional and illocutionary characteristics
   Time or length of response

---

5 RELATIONSHIP BETWEEN INPUT AND RESPONSE

Reciprocal
Nonreciprocal
Adaptive

Appendix II: WBPT – Test Taker Data

WBPT Section One – Version One – Test Taker Scores

ID	points out of 25	per cent
68	18	72
70	21	84
71	21	84
73	18	72
77	19	76
84	22	88
86	20	80
89	17	68
94	20	80
100	18	72
129	16	64
130	20	80
133	19	76
134	19	76
139	21	84
145	19	76
149	21	84
151	23	92
154	20	80
160	18	72
167	21	84
186	24	96
196	15	60
208	20	80
212	15	60
226	19	76
227	21	84
231	20	80
237	22	88
247	20	80
251	22	88
259	19	76
273	22	88
285	16	64
293	24	96
312	21	84
317	23	92
325	17	68
329	20	80
335	18	72
364	18	72
369	18	72
374	19	76

WBPT Section One – Version One – Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
19.6	78.5	2.2

WBPT Section One – Version Two – Test Taker Scores

ID	points out of 25	per cent
10	17	68
13	22	88
72	22	88
76	13	52
97	14	56
112	16	64
161	23	92
165	23	92
169	12	48
181	19	76
210	13	52
216	20	80
220	18	72
225	16	64
248	19	76
265	18	72
266	19	76
274	22	88
279	23	92
290	20	80
298	16	64
322	11	44
324	13	52
328	17	68
330	18	72
331	14	56
338	17	68
339	16	64
347	14	56
359	18	72
362	13	52
368	19	76
373	12	48

WBPT Section One – Version Two – Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
17.2	68.7	3.5

WBPT Section One – Version Three – Test Taker Scores

ID	points out of 25	per cent
12	18	72
75	21	84
81	16	64
99	13	52
104	14	56
107	21	84
113	15	60
114	16	64
143	13	52
183	12	48
194	15	60
206	14	56
213	15	60
214	15	60
217	17	68
224	17	68
240	17	68
272	10	40
275	7	28
287	17	68
294	11	44
296	19	76
304	21	84
307	16	64
326	13	52
351	21	84
358	18	72
363	13	52
365	14	56
370	14	56
371	16	64
372	17	68

WBPT Section One – Version Three– Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
15.5	62.0	3.2

WBPT Section One – Version Four – Test Taker Scores

ID	points out of 25	per cent
11	24	96
15	22	88
85	23	92
88	19	76
90	20	80
96	21	84
101	20	80
110	13	52
118	18	72
157	17	68
170	19	76
182	21	84
185	22	88
187	20	80
190	16	64
191	22	88
197	16	64
204	16	64
235	23	92
242	16	64
243	15	60
252	18	72
255	14	56
268	18	72
269	18	72
297	20	80
301	21	84
305	22	88
316	21	84
320	14	56
321	20	80
327	15	60
360	22	88
361	15	60

WBPT Section One – Version Four – Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
18.9	75.4	3.0

WBPT Section Two – Version One – Test Taker Scores

ID	points out of 25	per cent
10	14	56
68	22	88
77	23	92
100	18	72
154	21	84
161	25	100
163	24	96
181	18	72
183	19	76
204	13	52
210	13	52

	For how long have you been studying/learning English so far? Please type in the number of years.		
	Have you taken the Placement Test (Chemnitz University of Technology) before?	○ yes, when: ○ no	

227	24	96
231	23	92
237	23	92
255	12	48
305	25	100
312	25	100
316	16	64
324	14	56
330	22	88

WBPT Section Two – Version One – Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
19.8	79.2	4.4

WBPT Section Two – Version Two– Test Taker Scores

ID	points out of 25	per cent
76	18	72
111	20	80
114	23	92
119	20	80

139

Ancient Athens in new light

In Athens, the Parthenon and other ancient monuments on the Acropolis Hill were bathed in honey-colored light.

This is part of a project to highlight the city's monuments during the Olympics. The li[____] around the oth[____] high points o[____] the city we[____] dimmed a[____] hund[____] of spotlights aro[____] the Acropolis were swit[____] on. Fre[____] engineer Pierre Bideau, wh[____] designed the millennium lights-and-fireworks spectacle a[____] the Eiffel Tower, i[____] leading the Athens proj[____]. It will include nearby ancient sites whi[____] are linked b[____] a pedestrian walkway.

Athens is keen t[____] present its anci[____] heritage dur[____] the Olympics. ha[____] also launched a num[____] of programmes to tidy u[____] the city which has four million inhabi[____]. But much of the work o[____] ancient monuments wi[____] not be finished b[____] the begin[____] the Olympics.

Bideau and Micheal Cacoyannis, director of the 1964 movie "Zorba the Greek", work[____] together on the project which is supported by the culture ministry, the City of Athens and private donors.

http://sportsillustrated.cnn.com/2004/olympics/2004/07/19/bc.eu.spt.oly.athens.parthenon.ap cnn=yes, adapted by K.U.

143	20	80
167		

Placement Test - Microsoft Internet Explorer

Datei Bearbeiten Ansicht Favoriten Extras ?

Zurück Vorwärts Abbrechen Aktualisieren Startseite Suchen Favoriten Medien Verlauf

Adresse http://www.tu-chemnitz.de/phil/english/chairs/practlang/placement/backend/anzeige_ctest.php3

Canon Easy-WebPrint ▾ | Drucken Schnelldruck Vorschau | Optionen

Early UK springs become 'normal'

Many Britons find it acceptable that springs arrive early, scientists say.

They believe people ar[____] more and more us[____] to seeing spring flow
[____] at the New Year, and oth[____] signs of warm weat[____] several week
before usu[____]. They think the ear[____] arrival of spring sho[____] Nature
growing confusion. Th[____] Woodland Trust and th[____] Association for the
Advancement of Science are
runn[____] a project cal[____] Spring into Science.

The proj[____] asks people acr[____] the UK to rec[____] the timing o
[____] natural events and ha[____] found out th[____] many plants and anim
[____] can be seen abnor[____] early in the ye[____]. Jill Attenborough of the
Woodland Trust sa[____] people are now us[____] to seeing spring flow[____]
January, and th[____] year, for the first time ever, Scotland saw a bee in February
Sir David King, government scientist, says it is important that the public takes an
interest in the environment, and that people get outdoors and see for themselves ho
nature works.

http://news.bbc.co.uk/go/pr/fr/-/2/hi/science/nature/3502772.stm, published: 2004-3-21, ada
by K.U.

Fertig

173	23	92
187	17	68

(5)
Travel agent: I <u>do</u> apologise. I'd like to offer you a 20% <u>discount</u> on the price of one
our autumn breaks as a gesture of goodwill.

Tracy: A 20% discount, you must <u>joke</u>. I <u>want</u> to see the manager.

A	○	do
B	○	discount
C	○	joke
D	○	want

http://www.twinisles.com/english/edcmp.html, adapted by K.U.

Fertig

Tracy: <u>I'd like</u> to make a complaint <u>over</u> my holiday <u>in Portugal</u> last week.

A	○	can
B	○	I'd like
C	○	over
D	○	in Portugal

(2)
Travel agent: I'm sorry to hear that. What exactly was the problem?

Tracy: First of all the coach <u>taking</u> us to the hotel broke down and we had to wait f
over two hours in the <u>sweltering</u> heat before a replacement arrived. Then when we
to the hotel we found our room <u>hasn't been</u> cleaned.

A	○	taking
B	○	sweltering
C	○	got
D	○	hasn't been

(3)
Travel agent: Oh dear, did you complain to the hotel staff?

Tracy: Of course, but we were told all the chambermaids <u>are</u> off duty. Anyway, tha
not all. The people in the room <u>above</u> sounded <u>like</u> they were having all-night part
every night. I demanded another room but the <u>receptionist</u> told me the hotel was f

A	○	are
B	○	above
C	○	like
D	○	receptionist

(4) Travel agent: Oh, I <u>sea</u> .

Tracy: And to cap it all the food in the hotel restaurant <u>was</u> awful. It was so bad
had to <u>eat out</u> all the time <u>despite</u> having paid for meals in the price for our holid

A	○	sea
B	○	was
C	○	eat out
D	○	despite

191	23	92
194	16	64
197	15	60
213	19	76
216	22	88
224	23	92
226	21	84
251	19	76
265	18	72
273	18	72
285	19	76
304	22	88
325	18	72
327	21	84
328	22	88
359	23	92

WBPT Section Two – Version Two - Mean Scores and Standard Deviation

mean score (points out of 25)	mean score (per cent)	standard deviation
20.2	80.6	2.3

WBPT Section Two – Version Three – Test Taker Scores

ID	points out of 25	per cent
11	21	84
13	23	92
85	22	88
90	23	92
94	23	92
101	23	92
118	23	92
149	21	84
150	20	80
151	21	84
206	21	84
208	24	96
235	24	96
240	23	92
243	17	68
247	25	100
248	25	100
252	20	80
259	23	92
269	23	92
272	19	76
287	23	92
298	20	80
301	20	80
317	23	92
321	22	88
329	22	88
335	23	92
339	22	88
362	22	88

WBPT Section Two – Version Three – Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
22.0	88.1	1.7

WBPT Section Two – Version Four – Test Taker Scores

ID	points out of 25	per cent
71	20	80
74	20	80
75	22	88
84	19	76
130	24	96
165	20	80
254	23	92
274	21	84
277	23	92
279	24	96
290	21	84
297	20	80
307	18	72
314	21	84
351	23	92
358	16	64
360	22	88

WBPT Section Two – Version One – Mean Scores and Standard Deviation

mean score (points)	mean score (per cent)	standard deviation
21.0	84.0	2.1

Appendix III – TUC-PT – Test Taker Data

TUC-PT Section 2.1 – Test Taker Scores

ID	points out of 10	per cent
1	6	60
2	8	80
3	8	80
4	4	40
5	4	40
6	4	40
7	6	60
8	6	60
9	5	50
10	6	60
11	6	60
12	6	60
13	5	50
14	8	80
15	9	90
16	5	50
17	4	40
18	5	50
19	4	40
20	5	50
21	5	50
22	4	40
23	6	60
24	10	100
25	5	50
26	6	60
27	7	70
28	6	60
29	6	60
30	4	40
31	4	40
32	3	30
33	5	50
34	5	50
35	7	70
36	5	50
37	4	40
38	8	80
39	6	60
40	5	50
41	5	50
42	7	70
43	5	50
44	4	40
45	5	50
46	4	40

ID	points out of 10	per cent
47	7	70
48	2	20
49	5	50
50	7	70
51	3	30
52	5	50
53	8	80
54	4	40
55	4	40
56	3	30
57	6	60
58	6	60
59	4	40
60	5	50
61	5	50
62	8	80
63	5	50
64	4	40
65	3	30
66	6	60
67	7	70
68	5	50
69	7	70
70	6	60
71	8	80
72	4	40
73	5	50
74	1	10
75	5	50
76	9	90
77	7	70
78	4	40
79	7	70
80	5	50
81	3	30
82	5	50
83	3	30
84	5	50
85	5	50
86	6	60
87	3	30
88	4	40
89	5	50
90	4	40
91	6	60

ID	points out of 10	per cent
92	7	70
93	5	50
94	5	50
95	7	70
96	9	90
97	6	60
98	3	30
99	8	80
100	6	60
101	8	80
102	4	40
103	8	80
104	4	40
105	6	60
106	2	20
107	4	40
108	5	50
109	4	40
110	6	60
111	4	40
112	7	70
113	8	80
114	6	60
115	6	60
116	5	50
117	6	60
118	4	40
119	3	30
120	3	30
121	4	40
122	5	50
123	4	40
124	5	50
125	4	40
126	5	50
127	8	80
128	5	50
129	5	50
130	5	50
131	7	70
132	2	20
133	5	50
134	5	50
135	4	40
136	6	60
137	6	60

TUC-PT Section 2.1 - Mean Scores and Standard Deviation

mean score (points out of 10)	mean score (per cent)	standard deviation
5.3	53.1	1.6

TUC-PT Section 2.2 – Test Taker Scores

ID	points out of 14	per cent
1	9	64
2	12	86
3	10	71
4	11	79
5	9	64
6	12	86
7	11	79
8	12	86
9	10	71
10	10	71
11	8	57
12	8	57
13	10	71
14	11	79
15	8	57
16	11	79
17	13	93
18	8	57
19	6	43
20	8	57
21	10	71
22	10	71
23	12	86
24	11	79
25	11	79
26	11	79
27	14	100
28	12	86
29	10	71
30	9	64
31	7	50
32	10	71
33	9	64
34	11	79
35	10	71
36	11	79
37	9	64
38	11	79
39	10	71
40	11	79
41	10	71
42	10	71
43	8	57
44	11	79
45	7	50
46	8	57
47	9	64

ID	points out of 14	per cent
91	9	64
92	11	79
93	12	86
94	10	71
95	13	93
96	11	79
97	9	64
98	10	71
99	8	57
100	11	79
101	10	71
102	8	57
103	9	64
104	10	71
105	13	93
106	10	71
107	10	71
108	7	50
109	11	79
110	3	21
111	9	64
112	12	86
113	12	86
114	9	64
115	11	79
116	12	86
117	9	64
118	9	64
119	9	64
120	10	71
121	13	93
122	10	71
123	8	57
124	7	50
125	9	64
126	10	71
127	11	79
128	11	79
129	13	93
130	11	79
131	12	86
132	10	71
133	9	64
134	12	86
135	10	71
136	9	64
137	7	50

TUC-PT Section 2.2 - Mean Scores and Standard Deviation

Mean score (points)	Mean score (per cent)	Standard deviation
10.0	71.4	1.7

TUC-PT Section 2.3 – Test Taker Scores

ID	points out of 25	per cent
1	19	76
2	20	80
3	21	84
4	15	60
5	13	52
6	19	76
7	24	96
8	24	96
9	21	84
10	17	68
11	17	68
12	20	80
13	21	84
14	21	84
15	20	80
16	21	84
17	14	56
18	15	60
19	5	20
20	14	56
21	21	84
22	20	80
23	21	84
24	21	84
25	20	80
26	15	60
27	20	80
28	15	60
29	15	60
30	16	64
31	13	52
32	17	68
33	16	64
34	19	76
35	16	64
36	16	64
37	18	72
38	17	68
39	16	64
40	18	72
41	18	72
42	17	68
43	16	64
44	16	64
45	16	64
46	22	88

ID	points out of 25	per cent
47	17	68
48	12	48
49	17	68
50	18	72
51	15	60
52	18	72
53	23	92
54	21	84
55	11	44
56	13	52
57	14	56
58	18	72
59	14	56
60	16	64
61	18	72
62	21	84
63	21	84
64	19	76
65	16	64
66	16	64
67	19	76
68	19	76
69	15	60
70	19	76
71	17	68
72	11	44
73	17	68
74	8	32
75	20	80
76	16	64
77	15	60
78	15	60
79	17	68
80	16	64
81	17	68
82	20	80
83	21	84
84	23	92
85	19	76
86	19	76
87	14	56
88	13	52
89	13	52
90	14	56
91	19	76

ID	points out of 25	per cent
92	15	60
93	15	60
94	18	72
95	21	84
96	18	72
97	21	84
98	13	52
99	17	68
100	15	60
101	21	84
102	20	80
103	17	68
104	19	76
105	23	92
106	18	72
107	18	72
108	20	80
109	16	64
110	14	56
111	19	76
112	16	64
113	23	92
114	20	80
115	20	80
116	20	80
117	21	84
118	18	72
119	14	56
120	19	76
121	15	60
122	21	84
123	14	56
124	16	64
125	18	72
126	14	56
127	20	80
128	19	76
129	18	72
130	18	72
131	18	72
132	13	52
133	14	56
134	18	72
135	15	60
136	19	76
137	19	76

TUC-PT Section 2.3 - Mean Scores and Standard Deviation

Mean score (points)	Mean score (per cent)	Standard deviation
17.4	69.7	3.1

Appendix IV: Testing Materials (Base Texts)

"A British National Identity I"
WBPT Section One – Version One – Original Text

Snapshots of Britain
A sense of identity

The United Kingdom of Great Britain and Northern Ireland, to give it its formal title, is a highly centralised and unitary state, and its largest component, England, has been so for almost 1,000 years, longer than any other European country. By the sixteenth century Wales was fully incorporated into English administration and law but Britain as a political entity did not emerge until 1707, when the ancient kingdoms of Scotland and England were united. Ireland, which fell completely under English rule in the sixteenth century, became formally part of the United Kingdom in 1801 when, like Scotland a century earlier, it lost its own parliament. Ireland achieved independence in 1921, with the exception of six northern counties which remained part of the United Kingdom. Yet Northern Ireland is not part of Britain, although the term 'Britain' is often used loosely to mean the United Kingdom.

It is widely assumed that the British form a relatively homogeneous society with a strong sense of identity, but it is an assumption that requires considerable qualification. Even after 300 years the terms 'British' and 'Britain', which are used for official purposes, can seem very artificial. In his famous *Dictionary of Modern English Usage*, first published in 1925, Fowler wrote:

> *It must be remembered that no Englishman, or perhaps no Scotsman, calls himself a Briton without a sneaking sense of the ludicrous, or hears himself referred to as a BRITISHER without squirming. How should an Englishman utter the words* Great Britain *with the glow of emotion that for him goes with* England*? His Sovereign may be Her* Britannic *Majesty to outsiders, but to him is Queen of* England*.*

For centuries it has been the idea of England (or Scotland, or Wales), rather than Britain, which has been charged with patriotic emotion, particularly at times of national crisis. For example, at the Battle of Trafalgar in 1805, Admiral Lord Nelson's famous order to the British fleet read, 'England expects that every man will do his duty'. In 1939, during Parliament's emergency debate on the eve of war, one MP called across the chamber to another who was rising to speak: 'Come on, Arthur, speak for England.'

One should not be surprised, either that Fowler wrote the words quoted above under the entry for 'England'. If you look up 'Britain', 'British', and 'Briton', you will find 'See England'. Many people call Britain 'England', and the British 'English', as if Wales, Northern Ireland and Scotland were merely outer additions to England. Nothing, it should be said, infuriates the Scots, Welsh or Irish more than ignorantly to be called English, or for all Britain to be referred to as England. They have their own distinctive identities (see chapter 9).

Moreover, the Idea of England evokes images of the Queen, the Houses of Parliament, Westminster Abbey, the Tower of London and soft landscape of the southern counties of England. This is not surprising since almost one quarter of the British people live within 25 miles (40 km) of London's Trafalgar Square. But it also reveals that England as well as Britain is dominated by the south, and particularly the south east.

These popular images of England are very misleading. The United Kingdom is a land of great diversity, partly in its landscape, but more importantly in the human sphere. In addition to Scotland, Wales and Northern Ireland, the regions of England also have their special identities, which tend to be stronger the further one travels from London and the south east. In Cornwall, in the far south west, there is a reviving sense of Celtic identity, and a romantic affinity with their cousins, the people of Brittany in north-west France, persists. In the north of England, in the words of one MP, people are 'warm, friendly, quick-tempered and insular'. Communities in the north often have a strong sense of loyalty and identity. As one moves closer to London, community loyalties weaken and society is both more homogeneous and yet also more individualistic, the twin characteristics of a highly integrated modern society.

Each shire or 'county', the administrative divisions of England created over 1,000 years ago, still commands its own local loyalties, expressed in that most English of games, cricket. Even in the most homogeneous part of Britain, the 'Home Counties' around London (Middlesex, Hertfordshire, Essex, Kent, Surrey), people can still feel strongly about their county identity. The sense of local difference may be partly a matter of history, but it is also to do with the subtle changes in landscape, architecture or the way English is spoken.

England, unlike the largely mountainous countries of Wales and Scotland, is mainly lowland, except for six major hilly regions: the Pennines, called the 'backbone of England', dividing the north west of England from the north east; the scenic Lake District in the north west; the Yorkshire Dales, running to the east coast of Yorkshire; the moorlands of Cornwall and Devon in the south west; and the border areas with Scotland and Wales respectively. Elsewhere the ranges of hills are relatively low, while the East Midlands and East Anglia are notably flat and featureless. In Scotland and Wales the greater part of the population is concentrated in the more lowland areas, particularly the area between Glasgow and Edinburgh, and in the eastern and south-eastern parts of Wales.

(McDowall 1999: 10-12).

"A British National Idendity II"
WBPT Section One – Version One – Adapted Text

A British national identity (?)

The United Kingdom of Great Britain and Northern Ireland, to give it its formal title, is a highly centralised and unitary state.

Its largest part, England, (1) <u>has played</u> a leading role in the country for almost 1,000 years.

By the sixteenth century, Wales (2) <u>was</u> fully incorporated into English administration and law.

However, Britain as a (3) <u>political</u> entity did not emerge until 1707, when the ancient kingdoms of Scotland and England (4) <u>were united</u>.

Ireland, (5) <u>which</u> fell (6) <u>completely</u> under English rule in the sixteenth century, became formally part of the United Kingdom in 1801. (7) <u>Like</u> Scotland, it lost its own parliament.

Ireland achieved independence in 1921, with the (8) <u>exception</u> of six northern counties which remained part of the United Kingdom.

Yet Northern Ireland is not part of Britain, (9) <u>although</u> the term 'Britain' is often used loosely to mean the United Kingdom.

It is widely assumed that the British form a relatively homogeneous society with a strong sense (10) <u>of</u> identity.

But even today the terms 'British' and 'Britain', which (11) <u>are used</u> for official purposes, can (12) <u>seem</u> very artificial.

For centuries it has been the idea of England (and Scotland, or Wales), rather than of Britain, which (13) <u>has been charged</u> with patriotic emotion, particularly at times of (14) <u>national</u> crisis.

(15) <u>Many</u> people call Britain 'England', and the British 'the English', as if Wales, Northern Ireland and Scotland (16) <u>were</u> merely outer additions to England.

It infuriates the Scots, Welsh or Irish (17) <u>to be called</u> English, they have (18) <u>their</u> own distinctive identities.

Moreover, the idea of England evokes images of the Queen, the city of London and the soft landscape in the southern counties of England.

This is not so surprising since almost one quarter of the British (19) <u>population</u> live very close to London.

But it also reveals that England as well as Britain is dominated (20) <u>by</u> the south, and particularly the south east.

These popular images are very (21) <u>misleading</u>. The United Kingdom is a land of great diversity, partly in (22) <u>its</u> landscape, but (23) <u>more importantly</u> in the human sphere.

In addition to Scotland, Wales and Northern Ireland, the regions of England also (24) <u>have</u> their special identities, which tend (25) <u>to be</u> stronger the further one travels from London and the south east.

Mc Dowall (1999: 10ff.), adapted by K.U.

Brenda Fassie: A very human hero
Brenda Fassie - whom Time magazine dubbed the "Madonna of the Townships" - has died in South Africa aged 39.

She managed to combine ground-breaking musical success with an accessibility - and human fallibility - that drew a fierce loyalty and protectiveness from fans.

Her career was studded with record sales and awards, but punctuated also by periodic scandals, recurring battles with drug addiction, and lows in her musical career that saw her written off by the press.

Fassie's vocation seemed destined from the start. Born in Langa in the Cape Flats, she was named after US country singer Brenda Lee, and began to perform at five years old.

In her mid-teens, her talent was spotted when music producer Koloi Lebona heard her sing, and she headed for the big city of Johannesburg to begin her professional recording career.

'Something special'

"There was something special about her voice," Mr Lebona told South African news agency, Sapa.

"It was different to anything I had heard until then and was very mature for a 16 year old. I knew it was the voice of the future."

Fassie made good on her promise, shooting to fame in the early 1980s with the best-selling bubble-gum pop hit, Weekend Special, as Brenda and the Dudes.

It became the fastest selling record of the time and later moved into the international charts.

In the late 1980s, Fassie teamed up with record producer Sello "Chicco" Twala to launch Too Late for Mama - a multi-platinum seller. But then, Fassie found cocaine - and her fresh-faced success began to sour.

Rock bottom

A failed marriage, concert no-shows and a slide into debt followed.

In 1995, Fassie hit rock bottom. She was found in a hotel room, in a drug-induced haze - lying next to the body of her lesbian lover, Poppie Sihlahla, who had died of an apparent drug overdose.

> **I'd been shouting and shouting and no one wanted to hear me**
> Brenda Fassie on her album Memeza ('Shout')

Fassie booked into a drug rehabilitation centre.

In 1997, Fassie reunited with Mr Twala to record Memeza (Shout). She later said this album best epitomised her life.

"Memeza - it means 'Scream', 'Shout'," she said in 1999.

"I'd been shouting and shouting and no-one wanted to hear me. When I sing this song, I want to cry."

But, she said, the bad times were past.

New start?

"This is the beginning for me," she said.

It was a new start that confounded her critics. Memeza was South Africa's best-selling album of 1998 - and that same year, Fassie scooped the Kora Award for best female artist.

The best-selling wedding song Vulindlela, from the album, was used by the ruling African National Congress in its 1999 election campaign.

A string of best-selling albums followed, and for the next four consecutive years Fassie won a South African Music Award.

> **Brenda occupied a special place in the minds and hearts of many people around the world. Indeed, a hero has fallen**
> Leslie Sedibe
> EMI Music lawyer

She was on top of the world, saying: "I'm going to become the Pope next year. Nothing is impossible."

But Fassie never decisively won her battle with drugs and alcohol, and her visits to the rehab clinic continued.

Two weeks ago, she was rushed to hospital with breathing problems that led to an asthma attack and then cardiac arrest.

As her family sat in vigil around her bedside, fans besieged the clinic and Fassie was visited by former President Nelson Mandela, his former wife Winnie, and current President Thabo Mbeki.

On Sunday evening, she died, leaving behind an only son, Bongani.

"Death has this time robbed this country, Africa and the world of one of the greatest talents to come out of the township of Langa in Cape Town," said lawyer Leslie Sedibe from her record label EMI Music was quoted as saying by Sapa.

"Brenda occupied a special place in the minds and hearts of many people around the world. Indeed, a hero has fallen," he said.

(<http://news.bbc.co.uk/go/pr/fr/-/2/hi/africa/3700309.stm>, published: 2004/05/10)

"Brenda Fassie II"
WBPT Section One – Version Two – Adapted Text

Brenda Fassie – whom Time Magazine called the "Madonna of the Townships" – has died in South Africa (1) <u>aged</u> 39.

Brenda Fassie's career was full of record sales and awards, but her life was also punctuated (2) <u>by</u> periodic scandals, recurring battles with drug addiction and bad press coverage.

Fassie's vocation (3) <u>seemed</u> destined from the start. Brenda (4) <u>was named</u> after US country Brenda Lee, and began to perform (5) <u>at</u> five years old.

In her mid-teens, her talent was discovered, and her professional career began (6) <u>when</u> music producer Koloi Lebona heard (7) <u>her</u> sing, and she went to the big city of Johannesburg.

"There was something special about her voice," Mr Lebona told the South African news agency, Sapa.

"It was different to (8) <u>anything</u> I had heard (9) <u>until</u> then. I knew it (10) <u>was</u> the voice of the future."

Fassie and her band Brenda and the Dudes became famous in the early 1980s with the best-selling pop hit "Weekend Special".

It became (11) <u>the</u> fastest selling record of the time and later moved into the international charts.

In the late 1980s, Brenda Fassie got together with record producer Sello "Chicco" Twala, together they produced "Too Late For Mama" (12) <u>which</u> sold five million times.

But then Fassie(13) <u>became</u> addicted to cocaine – and her fresh success began to fade. Later, her marriage failed.

In 1995 she was(14) <u>found</u> druggy in a hotel room with the body of her lesbian lover, Poppie Sihlahla (15) <u>lying</u> next to her. Apparently, Poppie had died (16) <u>of</u> a drug overdose.

Brenda was taken to a drug rehabilitation centre.

In 1997, Fassie reunited with Sello Twala and they recorded the album "Memeza" (Shout).
The bad times were past.

She later said this album(17) <u>best summarized</u> her life.

It was a new start that surprised her (18) <u>critics</u>. "Memeza" was South Africa's best-selling album of 1998 – and that same year, Fassie received an award for best female artist.
A number of best-selling albums followed, and from 1999 to 2003 Brenda Fassie won a South African Music Award every year.

But Fassie never (19) <u>won</u> her battle with drugs and alcohol completely, (20) <u>although</u> she went into rehab clinics regularly.

(21) <u>Some time ago</u>, she was taken to hospital with breathing problems that led to an asthma attack and then cardiac arrest.

As her family sat around her bedside, fans were waiting (22) <u>in front of</u> the clinic and Brenda was visited by former president Nelson Mandela, his former wife Winnie, and (23) <u>current</u> president Thabo Mbeki.

On Sunday evening, she died leaving (24) <u>behind</u> a son, Bongani.

"eath has this time robbed this country, Africa and world (25) <u>of</u> one of the greatest to come out of the township of Langa in Cape Town," said lawyer Leslie Sedibe from her record label EMI Music. "Brenda occupied a special place in the minds and hearts of many people around the world. Indeed, a hero has fallen," he said.

(<http://news.bbc.co.uk/go/pr/fr/-/2/hi/africa/3700309.stm>, published: 2004/05/10, adapted by K.U.)

Teacher standards held for some schools
Schools in West benefit; Schools in South, not so much

ADEL, Oregon (AP) – There was a sigh of relief from Colorado to California after the Education Department announced that teachers in rural schools would have more time to meet strict federal qualifications.

But in the South, there was confusion. Although the region is home to hundreds of the country's most rural and poor areas, few schools there were granted the same reprieve.

"There are a lot of people that are very frustrated that for reasons that are arbitrary, their schools are not qualified," said Robert Lambert, a Kentucky-based researcher for the Rural Education and Community Trust, a nonprofit that studies rural issues. "There are schools that are small, isolated and poor, the they don't qualify."

Under federal requirements, all teachers must be "highly qualified" in every subject they teach, with a bachelor's degree or state certification in the topic.

The mandates are part of No Child Left Behind law, a centrepiece of the Bush administration's education policy.

Small school districts in the rural West and the Great Plains, where educators often teach several subjects to several different grades, had been struggling with the requirement until western lawmakers successfully lobbied on their behalf.

In March, it was announced that rural teachers would be allowed an extra year to prove that they met the "highly qualified" treshold, until 2007. New teachers would get three years from the date of their hire.

But outside the West and the Great Plains, far fewer schools will benefit from the changes.

That's because the federal government used criteria favoring small, self-contained districts like those in the West, instead of countywide districts like those in the South.

Collectively, that makes districts throughout the South too large to get the break extended to rural teachers, which the federal government made available only to schools that are enrolled in the Small Rural School Achievement program.

That program -- which gives extra money to districts with fewer than 600 students, in communities with fewer than 2,500 people -- serves about 5,000 schools, mostly in the West and the Midwest.

The upshot is that while 440 districts in Nebraska, 375 districts in Montana and 80 districts in Oregon qualified for the extra time, no districts in South Carolina or Alabama qualified. Only one district each qualified in Florida and West Virginia.

Janice Poda, who directs teacher quality for South Carolina's education department, said it came as a shot to learn that none of the schools in her state were being classified as rural.

"It has kind of become a joke," she said.

Doug Mesecar, deputy chief of staff for policy at the Education Department, said the agency set its criteria after meeting with educators in all 50 states. At those meetings, he said the changes seemed particularly pressing fro districts that were extremely isolated, with big cities three or more hours away.

"It is those districts that are particularly impacted by trying to get teachers qualified in multiple subjects," he said. "If we just said flexibility was available to anyone without parameters, it would be changing things dramatically."

But in a recent interview, Ray Simon, the assistant secretary for secondary and elementary education, said the criteria were being re-examined.

And Mesecar said, "If we have requests to look at providing this flexibility for a slightly larger district, we would be open to considering that, as long as it's not flexibility for its own sake, but flexibility where it is needed."

The change has been good news for Larry Ferguson, who teaches four subjects to five different grades at the tiny Adel middle school in south-central Oregon, often staying just two pages ahead of his students.

"You've got the same kids for five years, so you can't repeat anything -- you've got to come up with new stuff," Ferguson said. "It's as much an education for me as anything."

(<http://www.cnn.com/2004/EDUCATION/07/16/rural.school.ap/index.html>, published 2004/07/16)

"New Teacher Standards II"
WBPT Section One – Version Three – Adapted Text

New teacher standards in the United States
Schools in the west benefit, schools in the south not (1) **so much.**

There was a sigh of relief from Colorado to California after the Education Department announced that teachers in rural schools (2) <u>would have</u> more time to meet federal qualifications.

But in the south there was confusion. (3) <u>Although</u> the region is home to hundreds of the country's most rural and poor areas, (4) <u>few</u> schools were granted more time to meet the qualifications.

(5) <u>According to</u> federal requirements, all teachers must be "highly qualified" in every subject (6) _____ they teach, with a bachelor's degree or a state certification in the subjects.

These requirements are part of the No Child Left Behind law, an important chapter of the Bush administration's education policy.

In March, (7) <u>it</u> was announced that rural teachers would be allowed more time (8) <u>to achieve</u>
the required qualifications. New teachers would (9) <u>get</u> three years from the date of their hire.

But (10) <u>outside</u> the west and the Great Plains, far fewer schools will benefit (11) <u>from</u> these changes.

(12) <u>That's because</u> the federal government used criteria that are favoring small districts like those in the west, instead of larger districts (13) <u>like</u> those in the south.

The result is that, (14) <u>while</u> 440 districts in Nebraska, 375 districts in Montana and 80 districts in Oregon qualified for the extra time, no districts in South Carolina or Alabama
(15) <u>qualified</u>. Only one district each qualified in Florida and West Virginia.

Doug Mesecar, deputy chief of (16) <u>staff</u> for policy at the Education Department, said the agency (17) <u>set</u> its criteria after meeting with educators in all 50 states. At those meetings, he said the changes seemed (18) <u>particularly</u> urgent for districts that were extremely isolated.

"It (19) <u>is</u> those districts that are particularly impacted by trying to (20) <u>get</u> teachers qualified in several subjects," he said. "If we said anyone could get more time, this(21) <u>would be</u> changing things dramatically."

But in a recent interview, Ray Simon, the assistant secretary for secondary and elementary education, said the criteria (22) <u>were</u> being re-examined.

The change has been good news for Larry Ferguson, (23) <u>who</u> teaches four subjects to five different age groups at the tiny Adel middle school in south-central Oregon, often (24) <u>staying</u> just two pages ahead of his students.

"You've got the same kids (25) <u>for</u> five years, so you can't repeat anything – you've got to come up with new stuff."

(<http://www.cnn.com/2004/EDUCATION/07/16/rural.school.ap/index.html>, published 2004/07/16, adapted by K.U.)

Ancient Indians made 'rock music'
Archaeologists have rediscovered a huge rock art site in southern India where ancient people used boulders to make musical sounds in rituals.

The Kupgal Hill site includes rocks with unusual depressions that were designed to be struck with the purpose of making loud, musical ringing tones.

It was lost after its discovery in 1892, so this is the first fresh effort to describe the site in over a century.

Details of the research are outlined in the archaeological journal Antiquity.

A dyke on Kupgal Hill contains hundreds and perhaps thousands of rock art engravings, or petroglyphs, a large quantity of which date to the Neolithic, or late Stone Age (several thousand years BC).

Researchers think shamen or young males came to the site to carry out rituals and to "tap into" the power of the site. However, some of it is now at threat from quarrying activities.

Granite percussion

The boulders which have small, groove-like impressions are called "musical stones" by locals. When struck with small granite rocks, these impressions emit deep, "gong-like notes".

These boulders may have been an important part of formalised rituals by the people who came there.

In some cultures, percussion plays a role in rituals that are intended for shamen to communicate with the supernatural world. The Antiquity work's author, Dr Nicole Boivin, of the University of Cambridge, UK, thinks this could be the purpose of the Kupgal stones.

The first report of the site was in 1892, in the Asiatic Quarterly Review. But subsequent explorers who tried to find it were unable to do so.

Dr Boivin has been documenting the site. A few pictures of the site were taken in the 19th Century, but the originals were either lost, or allowed to fade.

Destruction imminent

Many of the motifs on the rocks are of cattle, in particular the long-horned humped-back type found in southern India (*Bos indicus*).

However, some are of human-like figures, either on their own or with cattle. Some of these in chains, or holding bows and arrows.

The typically masculine nature of the engravings leads Dr Boivin to suggest that the people who made the images were men and possibly those involved in herding cattle or stealing them.

The motifs themselves were made by bruising the rocks, presumably with a stone implement.

She believes that the people who made the motifs and those who went to view them must have been physically fit and agile.

Some of the images are in locations so difficult to reach that the artist must have suspended themselves - or got others to suspend them - from an overhang to make the images.

Modern-day commercial granite quarrying has already disturbed some sections of the hill. A rock shelter with even older rock art to the north of Kupgal Hill has been partially destroyed by quarrying.

"It is clear government intervention will be required to elicit effective protection for the majority of the sites in the [area] if these are not to be erased completely over the course of future years," writes Dr Boivin in Antiquity.

(<http://news.bbc.co.uk/go/pr/fr/-/1/hi/sci/tech/3520384.stm>, published: 2004/03/19)

Ancient Indians made 'rock music"
Archaeologists have rediscovered a huge rock art site in southern India where ancient people used pieces of rock to make musical sounds in rituals.

The Kupgal Hill site includes rocks with unusual furrows there (1) <u>were designed</u> a long time ago to be struck in order to make loud, musical ringing tones.

The Kupgal Hill site (2) <u>was discovered</u> in 1892. Soon after its discovery it was lost again. So this is the first effort to describe the site in over a century.

Details of the research can be found in the (3) <u>archaeological</u> journal *Antiquity*.

Kupgal Hill contains hundreds and perhaps (4) <u>thousands</u> of rock art engravings. A large number of them date (5) <u>to</u> the Neolithic, or late Stone Age (several hundred thousand years BC).

The rocks (6) <u>which</u> have small furrows are called "musical stones" by local (7) <u>people</u>. When (8) <u>struck</u> with small granite rocks, they emit deep sounds.

These rocks may have been an important part of formal rituals carried (9) <u>out</u> by the people who came there.

In some cultures, percussion plays a role in rituals that are used to (10) <u>communicate</u> with the supernatural world. The *Antiquity* author Dr Nicole Boivin, of the University of Cambridge, UK, thinks this could be the (11) <u>purpose</u> of the Kupgal stones.

The first report of the site was (12) <u>in</u> 1892, in the *Asiatic Quarterly Review*. But explorers who tried to find it afterwards were unable to (13) <u>do so</u>.

Dr Boivin has been documenting the site. (14) <u>A few</u> photographs of it were (15) <u>taken</u> in the 19th century, but most of the originals were lost.

Most of the motifs on the rocks are of humans, either on (16) <u>their own</u> or with cattle.

The fact that most of the figures are male leads Dr Boivin to suggest (17) <u>that</u> the images were made by men who were involved (18) <u>in</u> herding cattle or stealing them.

She believes that the people who made the motifs and (19) <u>those</u> who went to see them (20) <u>must have been</u> physically fit and agile.

The motifs were cut into the rocks, (21) <u>presumably</u> with a stone instrument.

Modern-day commercial granite mining has (22) <u>already</u> ruined some sections of the hill. One part of Kupgal has been partially destroyed by mining.

Dr Boivin writes in *Antiquity* that the government (23) <u>will have to protect</u> the majority of the sites in the area (24) <u>if</u> these are not (25) <u>to be destroyed</u> completely over the future years.

(<http://news.bbc.co.uk/go/pr/fr/-/1/hi/sci/tech/3520384.stm>, Published: 2004/03/19, adapted by K.U.)

Jane Austen – the hottest writer in Hollywood

When the BBC screened its latest adaptation of Jane Austen's novel *Pride and Prejudice* it was watched by a record 18 million British viewers. The series was then sold to 18 countries round the world, from America to Australia, from Iceland to Israel. There are Jane Austen's fans in all corners of the globe, and even special Jane Austen discussion groups on the Internet.

Jane Austen herself couldn't possibly have imagined this kind of worldwide fame. In her lifetime she never once travelled abroad, indeed she hardly ever left the south of England. When she died spinster in 1817, only four of her six novels had been published, all anonymously, and she had earned a grand total of £ 648.65 from her books. Now, nearly 200 years later, sales of her novels rival modern bestsellers such as John le Carré, reaching 35,000 a week. There have been film and television productions of not only *Pride and Prejudice*, but also *Emma*, *Persuasion* and the Oscar-winning *Sense and Sensibility*. Her house in Chawton in Hampshire is visited by 200 people a day.

The secret of her success

What makes her worldwide success so surprising is the narrowness of the world her stories portray, 'three or four families in a country village,' as Jane Austen herself said. However, according to Nigel Nicolson, author of *The World of Jane Austen*, the explanation for her enduring success is very simple: 'Her novels are love stories, always ending in a wedding. They show a wonderful understanding of the little moves that young people made then, and still do make towards and away from each other. They are also very funny.' Or, as the author P D James wrote, 'All the books have the same basic plot – searching for and finding the right mate.' Just as the most romantic novels, you may say, but difference is that these were written by a genius.

The life and loves of Jane Austen

She was born in 1775, the seventh of eight children. Her father was the Reverend George Austen. They were not well off, and lived in a rambling rectory in the village of Steventon in the Hampshire hills. The family often had to entertain themselves at home. By the time she was 12, Jane was writing stories about heroines imprisoned in haunted castles, being rescued by glamorous heroes.

In Jane's own life there were three romantic attachments. The first was a handsome Irish law student called Tom Lefroy, who she met in 1795, but who had returned to Ireland a year later. The second, in 1801 was a young man called Samuel Blackall, who she fell in love with when on holiday in Devon, but who tragically died suddenly, soon after. The third was a large young man called Harris Bigg-Wither, whose proposal she briefly accepted in 1802, but 'he had nothing to recommend him but his size', so she changed her mind.

By this time she was 27, and by the standards of the day, 'on the shelf'. She knew only too well that marriage was important for someone in her position, for the only work suitable for a penniless clergyman's was school teaching or being a governess. Jane wrote to her niece: 'Single women have a dreadful propensity for being poor – which is a very strong argument in favour of matrimony.' Thus in her novels it is not just love, but also money which makes the institution of marriage so important.

In 1801 the family had moved to Bath, where she was very unhappy. To make matters worse, in 1805 her father died, leaving his widow, Jane and her only sister Cassandra, also unmarried, even poorer than before. For four years they had to move from house to house, often staying with relatives. Jane must have felt particularly miserable at this time because she found it difficult to continue with her writing. Finally in 1809 her brother Edward allowed them to live in a house on his estate in Chawton, only a few miles from Steventon where she had grown up. Here she was much happier, despite being the poor relation, dependent on charity. She not only revised her earlier novels but was able to write new ones, using her experiences to satirize and make fun of the social inequalities she saw around her. At last in 1811, *Sense and Sensibility*, which may have been started as early as 1793, was the first of her novels to be published.

In 1816, Jane Austen fell ill with a disease of the kidneys. She died on July 18, 1817 in the arms of her sister, Cassandra. She was only 41.

Jane Austen, Hollywood star

The influence of cinema and television has led to worldwide fame for the quiet-living spinster with a sense of fun. People see the movie and then read the book. Not everyone is pleased by this. Winifred Wilson, member of the Jane Austen Society, says, 'These screen adaptations should have kept closer to the text. They are too heavy on romance and too light on satire. They shouldn't have been written with the sole aim of commercial success.' However, the actress Emma Thompson, who adapted *Sense and Sensibility* for the cinema, won't accept this. She says her screenplay is full of satire, and deals with the relationship between love and money. She went to Jane Austen's grave in Winchester Cathedral to say thank you for the Hollywood Oscar she won for the film. As she said at the Oscar ceremony in Los Angeles, 'I do hope Jane knows how big she is in Uruguay.'

(Soars & Soars 1998: 82-83)

"Jane Austen II"
WBPT Section Two – Version One – Adapted Text

Jane Austen – The hottest writer in Hollywood

When the BBC screened its latest adaptation of Jane Austen's *Pride and Prejudice*, it was watched by a record 18 million British viewers. The series was (1) th<u>en</u> sold to 18 (2) count<u>ries</u> around the world, from America to Australia, from Iceland to Israel.

There are Jane Austen's (3) fa<u>ns</u> in all corners (4) o<u>f</u> the globe, and (5) ev<u>en</u> Jane Austen (6) discu<u>ssion</u> groups on the Internet. Jane Austen herself couldn't (7) poss<u>ibly</u> have (8) imag<u>ined</u> this kind of worldwide (9) fa<u>me</u>. In her lifetime (10) sh<u>e</u> never once travelled (11) abr<u>oad</u>. When she died spinster (12) i<u>n</u> 1817, only four of (13) he<u>r</u> six novels had (14) be<u>en</u> published, all anonymously, and she (15) ha<u>d</u> earned a grand (16) tot<u>al</u> of £ 648.65 from her (17) boo<u>ks</u>.

Now, nearly 200 (18) yea<u>rs</u> later, sales (19) o<u>f</u> her novels rival (20) mod<u>ern</u> bestsellers, reaching 35,000 a (21) we<u>ek</u>. Today, there have (22) be<u>en</u> film and television (23) produc<u>tions</u> of not (24) on<u>ly</u> *Pride and Prejudice*, but (25) al<u>so</u> *Emma*, *Persuasion*, and the Oscar-winning *Sense and Sensibility*.

Her house in Chawton in Hampshire is visited by 200 people a day.

(Soars & Soars 1998: 82-83, adapted by K.U.)

"Ancient Athens I"

WBPT Section Two – Version Two – Original Text

Ancient Athens caught in new light for Olympics

ATHENS, Greece (AP) -- Athenians got a new reason to look up Monday.

The Parthenon and other ancient monuments on the Acropolis Hill were bathed in honeycomb-colored light as part of project to highlight the city's monuments before the Aug. 13-29 Olympics.

Illumination around other high points of the city was dimmed as hundreds of spotlights around the Acropolis -- known to Greeks as the "sacred rock" -- were switched on.

French engineer Pierre Bideau, who designed the millennium lights-and-fireworks spectacle at the Eiffel Tower, is heading the Athens venture. It will eventually include nearby ancient sites which are linked by a pedestrian walkway, as well as the city's Lycabettus Hill.

Athens is keen to show off its ancient heritage during the games, and has also launched a barrage of schemes to tidy up this crowded city of 4 million residents -- granting cash incentives to paint building facades while covering others with giant colorful banners.

But much of the improvement work on ancient monuments will not be finished by mid-August and scaffolding propping up parts of the 2,500-year-old Parthenon will still be there during the Olympics.

Bideau teamed up on the lighting project with Michael Cacoyannis, director of the 1964 movie "Zorba the Greek," on the project which has received backing from the culture ministry, the City of Athens and private donors.

"This is a great day for me and for all of Athens," Cacoyannis said. "We will work tirelessly to change the nighttime view of the city."

(<http://sportsillustrated.cnn.com/2004/olympics/2004/07/19/bc.eu.spt.oly.athens.parthenon.ap/index.html?cnn=yes>, published 2004/07/19)

"Ancient Athens II"
WBPT Section Two – Version Two – Adapted Text

Ancient Athens in new light

In Athens, the Parthenon and other ancient monuments on the Acropolis Hill were bathed in honey-colored light.

This is part of a project to highlight the city's monuments during the Olympics. The (1) lights around the (2) other high points (3) of the city (4) were dimmed as (5) hundreds of spotlights (6) around the Acropolis were (7) switched on. (8) French engineer Pierre Bideau, (9) who designed the millennium lights-and-fireworks spectacle (10) at the Eiffer Tower, (11) is leading the Athens (12) project.

It will include nearby ancient sites (13) which are linked (14) by a pedestrian walkway.

Athens is keen (15) to present its (16) ancient heritage (17) during the Olympics. It (18) has also launched a (19) number of programmes to tidy (20) up the city which has four million (21) inhabitants.

But much of the work (22) on ancient monuments (23) will not be finished (24) by the (25) beginning of the Olympics.

Bideau and Michael Cacoyannis, director of the 1964 movie "Zorba the Greek", work together on the project which is supported by the culture ministry, the City of Athens and private donors.

(<http://sportsillustrated.cnn.com/2004/olympics/2004/07/19/bc.e u.spt.oly.athens.parthenon.ap/index.html?cnn=yes>, published 2004/07/19, adapted by K.U.)

"Country Profile United Kingdom I"
WBPT Section Two – Version Three – Original Text

Country profile: United Kingdom
The United Kingdom is made up of England, Wales, Scotland and Northern Ireland. It has a long history as a major player in international affairs and fulfils an important role in the EU, UN and Nato.

The economy - one of the largest in the world - is no longer manufacturing but services-based, with e-commerce of growing significance. The City of London is a global financial centre.

Many of the country's people have never been richer, but a recent international study says the UK has the second highest child poverty rate in the European Union.

The country has not yet adopted the euro currency and the debate continues over when, and indeed whether, it will do so. The government has said that a series of economic criteria must be met before a referendum on the issue is held.

After decades of violent conflict in Northern Ireland, the Good Friday agreement of 1998, which led to a new assembly with devolved powers, brought hope of a lasting peace. While there has been progress, there are continuing arguments over arms decommissioning and alleged paramilitary activity. Tension between the unionist and nationalist communities still simmers.

The assembly was suspended in autumn 2002 when a row erupted over allegations about IRA activities. Elections were held in late 2003 but the deadlock continues and the assembly has yet to reconvene.

In recent years the United Kingdom has made significant moves on devolution of powers to Scotland and Wales as well. The Scottish Parliament in Edinburgh and the National Assembly for Wales in Cardiff opened in 1999, and the possibility of devolution for the English regions has also been discussed.

The United Kingdom is home to a multicultural population, partly as the legacy of empire. There are more Asian restaurants in many towns than there are burger bars and a former government minister has described a popular variety of curry as "Britain's true national dish".

The UK has been at the forefront of youth culture since the heyday of the Beatles and Rolling Stones in the 1960s.

It has a rich literary heritage encompassing the works of Englishman William Shakespeare, Scot Robert Burns, Welshman Dylan Thomas and Northern Irishman Seamus Heaney. Traditional music has deep roots across the UK which has also produced classical composers from Henry Purcell in the Baroque period to Benjamin Britten in the 20th century.

Population: 59.2 million (UN, 2003) **Capital:** London **Major language:** English **Major religion:** Christianity **Life expectancy:** 76 years (men), 81 years (women) (UN) **Monetary unit:** 1 pound sterling = 100 pence **Main exports:** Manufactured goods, chemicals, foodstuffs **GNI per capita:** US $25,510 (World Bank, 2002) **Internet domain:** .uk **International dialling code:** +44

Head of state: Queen Elizabeth II

Prime minister: Tony Blair

Tony Blair became prime minister in May 1997 when he led the Labour Party back into government after 18 years in opposition. He won a second resounding electoral victory in June 2001.

Mr Blair defines his politics as "the third way". His social democracy recognises that free-market capitalism has virtues, but he says that it also emphasises the role of the state in developing social justice and equality.

Mr Blair has been an unflinching supporter of the US-led campaign against international terrorism. Following 11 September, British forces were involved in operations in Afghanistan.

As the focus shifted to Baghdad, Mr Blair remained energetic in his diplomatic efforts but was also firm in insisting that Iraq risked war unless it complied with UN Security Council Resolution 1441 relating to weapons of mass destruction.

In the end, Mr Blair's government joined that of the USA in launching military strikes against Iraq without UN approval and despite domestic public disquiet and unease within his own party.

As efforts to find weapons of mass destruction in Iraq following the overthrow of Saddam Hussein failed to discover any, there was controversy in the UK about the government's case for going to war.

The 2001 general election campaign highlighted a public perception that there is not enough government investment in health care and education. Its performance in these areas is likely to be the key issue at the next election.

Chancellor of the exchequer: Gordon Brown Foreign secretary: Jack Straw Home secretary: David Blunkett Defence secretary: Geoff Hoon

The UK has a strong tradition of public-service broadcasting and an international reputation for creative programme-making.

The fledgling BBC began daily radio broadcasts in 1922 and quickly came to play a pivotal role in national life. The Empire Service - the forerunner of the BBC World Service - established a reputation worldwide. The BBC is funded by a licence fee, which all households with a TV set must pay.

BBC Broadcasting House, a London landmark

Commercial TV began in 1955 with the launch of ITV. Commercial radio was introduced in the 1970s, although ship-based pirate radio stations flourished in the 1960s before being outlawed. Hundreds of privately-owned radio and TV stations now compete with the BBC for listeners and viewers.

Home-grown soap operas have long topped the TV ratings, and British viewers keenly follow the ups and downs of life in east London's Albert Square, the setting for the

191

BBC's EastEnders, and Coronation Street - ITV's soap about northern-English working-class life. Programmes which catapult ordinary people into the public eye - known as reality TV - are enjoying a wave of popularity.

The UK's main TV networks now face strong competition from digital satellite and cable TV, which offer scores of channels. Digital terrestrial TV carries a smaller number of free-to-view channels. Digital radio (DAB) has had a slower start, but the BBC and commercial operators provide digital-only radio services.

The British media are free and able to report on all aspects of British life. The variety of publications on sale reflects the full spectrum of political opinion, as well as the British public's voracious appetite for newspapers.

In 2004 the BBC was plunged into crisis over the outcome of a judge's inquiry into the suicide of a scientist involved in a row over a BBC radio report. The item claimed that the government had embellished its case for war in Iraq.

The judge, Lord Hutton, exonerated the government, said the most serious claims in the BBC report were unfounded and reached damaging conclusions about BBC processes.

The press
Daily Telegraph - broadsheet Financial Times - daily business broadsheet The Guardian - daily broadsheet The Independent - daily broadsheet The Times - daily broadsheet The Sun - daily tabloid The Mirror - daily tabloid The Scotsman - daily broadsheet The Daily Mail - tabloid The Daily Express - daily tabloid

Television
BBC TV - operates BBC1, BBC2 and digital services including BBC News 24 BBC World - commercially-funded international news channel ITV - major commercial network, organised around regional franchises Channel 4 - national commercial station Five - national entertainment-based channel Independent Television News (ITN) - supplier of news to ITV, Channel 4, Five British Sky Broadcasting (BSkyB) - operator of digital satellite TV platform and provider of film, entertainment channels and rolling news channel Sky News

Radio
BBC Radio - national services include new-music station Radio 1, adult music station Radio 2, cultural network Radio 3, flagship speech station Radio 4 and news and sport station Five Live BBC Radio Scotland BBC Radio nan Gaidheal - Gaelic-language station for Scotland BBC Radio Ulster - for Northern Ireland BBC Radio Wales BBC Radio Cymru - Welsh-language BBC Asian Network - for Asian communities in the UK BBC World Service - major international broadcaster, heard worldwide via shortwave and increasingly on FM relays, programmes in more than 40 languages Virgin Radio - national commercial pop and rock station Talk Sport - national commercial sports station Classic FM - national commercial classical music station

News agency
The Press Association

(<http://news.bbc.co.uk/go/pr/fr/-/2/hi/europe/country_profiles/1038758.stm>, published: 2004/02/29)

Country Profile: United Kingdom

The United Kingdom consists of England, Wales, Scotland, and Northern Ireland. (1) I<u>t</u> has a (2) lo<u>ng</u> history as a (3) maj<u>or</u> player in (4) interna<u>tional</u> affairs, and (5) pla<u>ys</u> an important role in the EU, UN and Nato.

The economy – (6) on<u>e</u> of the (7) larg<u>est</u> in the world – (8) i<u>s</u> no (9) long<u>er</u> manufacturing- but services-based, (10) an<u>d</u> e-commerce is (11) beco<u>ming</u> more and more important.

(12) Th<u>e</u> City of London (13) i<u>s</u> a global (14) finan<u>cial</u> centre. The (15) coun<u>try</u> has not (16) ye<u>t</u> adopted the euro (17) curr<u>ency</u>. There are (18) sti<u>ll</u> debates (19) ov<u>er</u> when and (20) whe<u>ther</u> it will do (21) s<u>o</u>.

The United Kingdom (22) ha<u>s</u> a (23) multicu<u>ltural</u> population, (24) whi<u>ch</u> is partly a historical result (25) o<u>f</u> the British empire.

(<http://news.bbc.co.uk/go/pr/fr/-/2/hi/europe/country_profiles/1038758.stm>, published: 2004/02/29, adapted by K.U.)

"Early UK Springs I"
WBPT Section Two – Version Four – Original Text

Early UK springs become 'normal'
The effects of climate change mean there is increasing acceptance by Britons of spring arriving early, scientists say.

They believe people are more and more accustomed to seeing spring flowers at the New Year, and other signs of warm weather several weeks before usual.

Launching National Science Week, they said the premature arrival of spring displayed Nature's growing confusion.

They also urged scientists to reach out to people and invite them to discuss the issues which were worrying them.

The Woodland Trust and the British Association for the Advancement of Science are together running one project linked to the week, Spring into Science.

Midwinter visitor

This asks people across the UK to join in recording the timing of natural events, known as phenology, and has found that there are many abnormally early sightings of plants and wildlife.

> **For the sake of our future, it is important that we encourage the public to take an interest in the environment**
> Sir David King

Jill Attenborough of the trust said: "The warm winter weather, until mid-February, encouraged early observations of many species, but recent colder weather has slowed the advance of spring.

"We anticipate that warming weather will bring a rapid rush of activity. Bumble bees will become much more frequent visitors to gardens."

The project says the earliest recorded sighting of a bumble bee this season was shortly before Christmas near London. In recent years, the first sightings of the insects were on average two to three weeks earlier than 25 years ago.

Future investment

Jill Attenborough told BBC News Online: "This would have been very unusual 30, 40 or 50 years ago. Now, year on year, we are beginning to accept it as the norm.

"People are used to seeing daffodils in January, and this year Scotland had the first ever sighting of a bumble bee in February."

Sir David King, the government's chief scientific adviser, said National Science Week mattered: "For the sake of our future, it is important that we encourage the public to take an interest in the environment, to get outdoors and see for ourselves how nature works."

The president of the BA, Dame Julia Higgins, said: "Of course it is important that during National Science Week we inspire and entertain visitors at science events.

"But it is equally important that the scientists take this opportunity to reach out to the public and give them the opportunity to discuss the issues which affect and concern them."

(<http://news.bbc.co.uk/go/pr/fr/-/2/hi/science/nature/3502772.stm>,

published: 2004/03/12)

Early UK springs become 'normal'
Many Britons find it acceptable that springs arrive early, scientists say.

They believe people (1) ar<u>e</u> more and more (2) us<u>ed</u> to seeing spring (3) flow<u>ers</u> at the New Year, and (4) oth<u>er</u> signs of warm (5) wea<u>ther</u> several weeks before (6) usu<u>al</u>.

They think the (7) ear<u>ly</u> arrival of spring (8) sho<u>ws</u> Nature's growing confusion.

(9) Th<u>e</u> Woodland Trust and (10) th<u>e</u> British Association for the Advancement of Science are (11) run<u>ning</u> a project (12) cal<u>led</u> Spring into Science.

The (13) pro<u>ject</u> asks people (14) ac<u>ross</u> the UK to (15) rec<u>ord</u> the timing (16) o<u>f</u> natural events and (17) ha<u>s</u> found out (18) th<u>at</u> many plants and (19) anim<u>als</u> can be seen

(20) abnor<u>mally</u> early in the (21) ye<u>ar</u>.

Jill Attenborough of the Woodland Trust (22) sa<u>ys</u> people are now (23) us<u>ed</u> to seeing sping (24) flow<u>ers</u> in January, and (25) th<u>is</u> year, for the first time ever, Scotland saw a bee in February.

Sir David King, government scientist, says it is important that the public takes an interest in the environment, and that people get outdoors and see for themselves how nature works.

(<http://news.bbc.co.uk/go/pr/fr/-/2/hi/science/nature/3502772.stm>, adapted by K.U.)

Appendix V – Error Rates for Items Discussed

1. WBPT Items

[5.4.1.1.I] (WBPT Section One – Version One – Item One)

[5.4.1.1.I]

The United Kingdom of Great Britain and Northern Ireland, to give it its formal

title, is a highly centralised and unitary state. Its largest part, England

_____

4) <u>has played</u>

5) is playing

6) plays

a leading role in the country for almost 1,000 years.

ID	incorrect alternative chosen
70	2
71	2
73	3
100	3
130	2
145	3
149	2
167	3
196	3
212	3
227	3
312	3
335	2
364	2
369	3
374	3

[5.4.1.1.I] – Error Rates
(Percentages of the Total Group of Test Takers for Version One)

alternative 2	alternative 3	total
14%	23%	37%

(total number of test takers = 43)

[5.4.1.1.II]

Brenda _____

4) had been named

5) has been named

6) <u>was named</u>

after US country singer Brenda Lee

ID	incorrect alternative chosen
10	2
76	2
112	2
169	1
210	2
248	1
265	2
274	1
298	2
322	1
324	2
328	2
362	2

[5.4.1.1.II] –Error Rates
(Percentage of Total Group of Test Takers for Version Two)

alternative 1	alternative 2	total
12%	27%	37%

(total number of test takers = 33)

[5.4.1.1.V]

The result is that (14) <u>while</u> 400 districts in Nebraska, 375 districts in Montana and 80 districts in Oregon qualified for the extra time, no districts in South Carolina or Alabama _____

4) had qualified.

5) <u>qualified.</u>

6) qualify.

ID	incorrect alternative chosen
12	1
75	1
81	1
99	1
104	1
114	1
183	1
206	1
214	1
272	1
275	1
287	3
294	1
307	1
326	1
358	1
365	1
370	1
371	1

[5.4.1.1.V] – Error Rates
(Percentage of Total Group of Test Takers for Version Three)

alternative 1	alternative 3	total
53%	3%	56%

(total number of test takers = 32)

[5.4.1.1.VI]

But in a recent interview, Ray Simon, the assistant secretary for secondary and elementary education, said the criteria _____

4) is

5) was

6) <u>were</u>

being re-examined.

ID	incorrect alternative chosen
12	2
75	2
81	1
99	2
104	2
107	2
113	2
114	2
143	2
183	2
213	1
214	2
217	1
240	1
272	1
275	2
287	1
294	2
296	1
304	2
307	2
326	1
358	1
363	2
365	2
370	2
372	2

alternative 1	alternative 2	total
28%	56%	84%

(total number of test takers = 32)

[5.4.1.1.VIII]

She believes that the people who made the motifs and (19) <u>those</u> who went to

see them _____

4) must be

5) <u>must have been</u>

6) will be

physically fit and agile.

ID	incorrect alternative chosen
197	1
204	1
242	1
243	1
255	1
269	1
320	3
361	1

[5.4.1.1.VIII] – Error Rates
(Percentages of the Total Group of Test Takers for Version Four)

alternative 1	alternative 3	total
21%	3%	24%

(total number of test takers = 34)

[5.4.1.2.I]

Brenda (4) <u>was named</u> after US country singer Brenda Lee and began to perform

_____

4) <u>at</u>

5) with

6) --

five years old.

ID	incorrect alternative chosen
10	2
13	3
76	2
97	3
112	2
169	3
181	2
216	2
248	3
265	3
266	3
274	2
290	3
324	2
328	3
331	2
359	3
373	2

[5.4.1.2.I] – Error Rates
(Percentages of the Total Group of Test Takers for Version Two)

alternative 2	alternative 3	total
27%	27%	55%

(total number of test takers = 33)

[5.4.1.2.VI]

On Sunday evening, she died, leaving _____

4) back

5) <u>behind</u>

6) out

a son, Bongani.

ID	incorrect alternative chosen
10	1
76	1
97	1
181	1
210	1
225	1
248	1
298	1
322	1
330	1
338	1
339	1
362	1
368	1
373	3

[5.4.1.2.VI] – Error Rates
(Percentages of the Total Group of Test Takers for Version Two)

alternative 1	alternative 3	total
42%	3%	45%

(total number of test takers = 33)

[5.4.1.2.VII] (WBPT Section One – Version Four – Item Nine)

[5.4.1.2.VII]

These rocks have been an important part of formal rituals carried

_____

1)   on

2)   <u>out</u>

3)   through

by the people who came there.

ID	incorrect answer chosen
15	1
85	3
90	1
96	1
101	1
110	1
118	1
157	1
170	1
190	1
191	1
197	1
204	1
242	1
243	1
268	1
297	1
301	1
320	1
321	1
327	3
360	3
361	1

[5.4.1.2.VII] – Error Rates
(Percentages of the Total Group of Test Takers for Version Four)

alternative 1	alternative 3	total
59%	9%	68%

(total number of test takers = 34)

[5.4.1.2.VIII]

But _____

1) out of

2) <u>outside</u>

3) outwards

the west and the Great Plains, far fewer schools will benefit (11) <u>from</u> these changes.

ID	incorrect alternative chosen
12	1
99	3
107	1
113	3
143	1
183	1
194	3
217	3
224	3
240	3
275	1
294	1
307	3
326	1
358	1
363	1
370	3
371	3
372	1

[5.4.1.2.VIII] – Error Rates
(Percentages of the Total Group of Test Takers for Version Three)

alternative 1	alternative 3	total
31%	28%	59%

(total number of test takers = 32)

[5.4.1.2.IX]

"You've got the same kids _____

1) <u>for</u>

2) since

3) up to

five years, so you can't repeat anything – you've got to come up with new stuff."

ID	incorrect alternative chosen
143	2
183	2
194	2
213	2
272	2
275	2
304	2

[5.4.1.2.IX] – Error Rates
(Percentages of the Total Group of Test Takers for Version Three)

alternative 2	alternative 3	total
22%	0%	22%

(total number of test takers = 32)

[5.5.1.1.I]

      Jane Austen couldn't (7) poss_____ (possibly) have **(8) imag_____**

      (imagined) this kind of worldwide (9) fa_____ (fame)

ID	incorrect letter(s) supplied
210	*i-n-e*
225	*i-n-e*
227	*i-n-e*
255	*i-n-e*
324	*i-n-e*

[5.5.1.1.I] – Error Rates
(Percentages of the Total Group of Test Takers for Version One)

i-n-e	total
24%	24%

(total number of test takers = 21)

[5.5.1.1.II]

..., and she **(15) ha**_____ (had) earned a grand (16) tot_____ (total) of £ 648.65 from her (17) boo_____ (books).

ID	incorrect letter(s) supplied
204	-*s*
210	-*s*
255	-*s*
316	-*s*
324	-*s*

[5.5.1.1.II] – Error Rates
(Percentages of the Total Group of Test Takers for Version One)

-*s*	total
24%	24%

(total number of test takers = 21)

[5.5.1.1.IV]

(8) Fre_____ (French) engineer Pierre Bideau, (9) wh_____ (who) designed the millennium lights-and-fireworks spectacle (10) a_____ (at) the Eiffel Tower, **(11) i_____** (is) leading the Athens (12) proj_____ (project).

ID	incorrect letter(s) supplied
111	*-n*
143	*-n*
187	*-n*
190	*-n*
194	*-t*
197	*-n*
325	*-n*
359	*-n*

[5.5.1.1.IV] – Error Rates
(Percentages of the Total Group of Test Takers for Version Two)

-n	*-t*	total
27%	4%	31%

(total number of test takers = 26)

[5.5.1.1.V] (WBPT Section Two – Version Three – Item Eleven)

[5.5.1.1.V]

The economy – (6) on_____ (one) of the (7) larg_____ (largest) in the world – (8) i_____ (is) no (9) lon_____ (longer) manufacturing- but services-based, (10) an_____ (and) e-commerce is **(11) beco**_____ (becoming) more and more important.

ID	incorrect letter(s) supplied
243	*-m-e*
272	*-m-e*
298	*-m-e*

[5.5.1.1. V] – Error Rates
(Percentages of the Total Group of Test Takers for Version Three)

-m-e	total
10%	10%

(total number of test takers = 30)

[5.5.1.2.I]

sales **(19) o_____** (of) her novels rival (20) mod_____ (modern) bestsellers.

ID	incorrect letter(s) supplied
10	*-v-e-r*
181	*-n-l-y*
183	*-n*
204	*-n*
210	*-v-e-r*
255	*-v-e-r*

[5.5.1.2.I] – Error Rates
(Percentages of the Total Group of Test Takers for Version One)

-v-e-r	*-n*	*-n-l-y*	total
14%	10%	5%	29%

(total number of test takers = 21)

[5.5.1.2.II]

The (13) proj_____ (project) asks people (14) acr_____ (across) the UK to (15) rec_____ (record) the timing **(16) o**_____ (of) natural events

ID	incorrect letter(s) supplied
290	*-n*
358	*-n*
376	*-v-e-r*

[5.5.1.2.II] – Error Rates
(Percentages of the Total Group of Test Takers for Version Four)

-n	*-v-e-r*	total
12%	6%	18%

(total number of test persons = 17)

[5.5.1.2.III]

The (1) lig_____ (lights) around the (2) oth_____ (other) high points **(3)**

o_____ (of) the city (4) we_____ dimmed

ID	incorrect letter(s) supplied
216	*-n*
273	*-v-e-r*
285	*-v-e-r*
304	*-v-e-r*

[5.5.1.2.III] – Error Rates
(Percentages of the Total Group of Test Takers for Version Two)

-v-e-r	*-n*	total
11%	4%	29%

(total number of test takers = 26)

[5.5.1.2.VI]

(8) Fre_____ (French) engineer Pierre Bideau, (9) wh_____ (who) designed the millennium lights-and-fireworks spectacle **(10) a_____** (at) the Eiffel Tower

ID	incorrect letter(s) supplied
76	*-r-o-u-n-d*
111	*-r-e*
119	*-r-o-u-n-d*
173	*-s*
187	*-n-d*
191	*-r-o-u-n-d*
194	*-s*
197	*-s*
251	*-s*
273	*-s*
285	*-n-d*
325	*-s*
327	*-r-o-u-n-d*

[5.5.1.2.VI] – Error Rates
(Percentages of the Total Group of Test Takers for Version Two)

-s	*-r-o-u-n-d*	*-n-d*	*-r-e*	total
23%	15%	8%	4%	50%

(total number of test takers = 26)

2. TUC-PT Items

[5.4.2.1.I] (TUC-PT Section 2.1 – Item Four)

[5.4.2.1.I]

Just look at the dog – he's completely covered in mud! What on earth . . . he . . .?

a. was . . . doing   b. had . . . done

c. did . . . do   d. has . . . doing

ID	incorrect alternative chosen
2	a
3	a
4	a
5	b
6	b
7	b
9	b
11	b
12	b
13	b
16	b
18	a
19	a
21	b
22	b
23	b
25	b
26	b
27	b
28	c
29	b
30	c
31	c
32	a
35	b
36	a
37	b
38	b
39	b
40	b
41	b
42	b
43	b
45	b
46	b

ID	incorrect alternative chosen
47	b
48	a
49	c
50	b
51	b
54	a
55	c
56	a
57	c
59	c
60	b
61	b
63	b
64	b
65	b
66	b
67	b
68	b
69	b
70	a
71	b
73	b
74	b
75	c
77	c
78	c
80	b
81	b
82	b
83	c
84	b
85	a
87	a
88	c
90	c
91	b
92	a
93	a
94	b
95	b
97	b
98	a
100	c
102	b
105	b
106	c

ID	incorrect alternative chosen
107	a
108	b
109	c
110	c
111	a
112	a
113	b
114	b
115	b
116	b
117	b
118	b
119	b
120	a
122	b
123	a
124	b
125	b
126	c
127	b
128	b
129	a
130	b
131	b
132	b
133	a
134	c
135	b

[5.4.2.1.I] - Error Rates
(Percentages of the Total Target Group)

alternative a	alternative b	alternative c	total
17%	49%	14%	80%

(total number of test takers = 137)

[5.4.2.1.II]

"Tom told me yesterday that he <u>. . .</u> to see us this morning, but it's already 2 o'clock and he still hasn't arrived."

a. comes   c. came

b. will come   <u>d. would come</u>

ID	incorrect alternative chosen
6	a
11	b
12	b
20	b
33	b
40	b
41	b
42	b
43	b
48	a
51	b
52	b
60	b
61	a
63	b
64	b
65	a
73	b
74	b
80	b
81	b
83	c
87	b
88	b
90	b
93	b
98	b
99	b
104	b

ID	incorrect alternative chosen
106	b
107	b
111	b
118	b
119	b
120	b
123	a
129	a
135	b

[5.4.2.1.II] – Error Rates
(Percentages of the Total Target Group)

alternative a	alternative b	alternative c	total
4%	23%	1%	28%

(total number of test takers = 137)

[5.4.2.2.I]

Please finish this test <u>. . .</u> lunchtime.

a. until   <u>c. by</u>

b. to   d. till

ID	incorrect alternative chosen
1	d
2	a
4	a
5	a
6	d
7	d
8	d
9	a
13	a
16	a
17	a
18	a
19	d
20	d
21	a
22	d
23	d
25	a
26	a
27	d
28	a
29	a
30	b
31	d
32	a
33	a
34	a
35	d
36	d
37	d
39	a
41	d
43	a
44	d
45	a

ID	incorrect alternative chosen
46	a
47	a
48	a
49	d
51	d
52	d
54	a
55	a
56	a
57	a
58	a
59	a
61	d
63	a
64	a
65	a
66	d
67	a
68	d
69	a
70	d
72	a
73	a
74	d
75	a
78	d
80	a
81	a
82	a
83	a
84	a
85	a
86	a
87	d
88	d
89	a
90	d
91	d
93	a
94	a
98	d
100	d
101	d
102	a
103	a
104	a
105	d

ID	incorrect alternative chosen
106	a
107	a
108	a
109	a
110	d
111	d
114	a
115	d
118	a
119	d
120	d
121	a
122	a
123	d
124	a
125	a
126	a
128	d
129	d
130	a
132	a
133	a
135	d

[5.4.2.2.I] – Error Rates
(Percentages of the Total Target Group)

alternative a	alternative b	alternative d	total
46%	1%	30%	77%

(total number of test takers = 137)

[5.4.2.2.II]

My rabbit is bigger ***like / <u>than</u> / as*** my cat.

ID	incorrect alternative chosen
12	3
19	3
31	1
99	3
103	3
110	3
111	3

[5.4.2.2.II – Error Rates
(Percentages of the Total Target Group)

alternative 1	alternative 3	total
1%	4%	5%

(total number of test takers = 137)

[5.5.2.1.I] (TUC-PT Section 2.3 – Item One)

[5.5.2.1.I]

Some time ago [1] I *was spending / <u>spent</u> / have spent* three weeks in Barcelona

ID	incorrect alternative chosen
1	3
2	1
5	3
6	3
10	1
13	1
18	1
19	1
20	1
21	1
28	3
32	1
34	1
35	3
38	3
43	3
45	1
53	1
55	3
57	3
58	3
66	1
71	3
72	3
73	3
74	3
76	3
78	3
81	3
86	3
87	1
88	3
89	3
103	1
104	1
106	3
109	3
110	3
112	3

ID	incorrect alternative chosen
122	3
126	3
128	3
129	3
133	3
134	3
135	3
136	3

[5.5.2.1.I] – Error Rates
(Percentages of the Total Target Group)

alternative 1	alternative 3	total
11%	23%	34%

(total number of test takers = 137)

[5.5.2.1.II]

A few days later I [12] *was receiving / had received / <u>received</u>* a call from some acquaintances

ID	incorrect alternative chosen
19	2
26	1
29	2
31	2
37	1
55	2
74	1
85	2
90	1
98	1
109	2
110	1
133	2
134	1
135	1

[5.5.2.1.II] – Error Rates
(Percentages of the Total Target Group)

alternative 1	alternative 2	total
6%	5%	11%

(total number of test takers = 137)

[5.5.2.1.III] (TUC-PT Section 2.3 – Item 23)

[5.5.2.1.III]

Now I wonder what changes [23] **have occurred / <u>will have occurred</u> / will occur** by the next time I have a chance to go to Barcelona

ID	incorrect alternative chosen
3	3
10	3
12	3
14	3
15	3
20	3
22	1
25	3
26	3
28	3
29	3
30	3
31	1
33	3
35	3
37	1
40	3
41	3
42	3
43	3
44	3
45	3
47	3
49	3
51	3
55	3
56	1
57	3
58	3
59	3
62	3
68	3
72	3
74	3
76	3
79	3
80	3
83	3
84	3

ID	incorrect alternative chosen
85	3
87	3
88	1
90	3
91	3
92	3
94	3
96	3
98	1
100	3
110	3
111	3
116	3
118	3
121	3
126	3
127	3
128	3
130	3
132	3
134	3
135	1
137	3

[5.5.2.1.III] – Error Rates
(Percentages of the Total Target Group)

alternative 1	alternative 3	total
5%	40%	45%

(total number of test takers = 137)

[5.5.2.2.I]

> so we returned to a place we knew [11] ***close to / in the near of / near of*** the Old Quarter.

ID	incorrect alternative chosen
2	2
5	2
6	2
10	3
18	3
19	2
23	3
26	3
34	2
35	2
39	2
47	2
52	2
57	2
60	2
61	2
65	2
66	2
69	3
72	2
73	2
74	2
77	3
93	3
95	2
96	2
103	2
110	3
121	2
124	2
136	2
137	2

[5.5.2.2.I] – Error Rates
(Percentages of the Total Target Group)

alternative 2	alternative 3	total
17%	6%	23%

(total number of test takers = 137)

Deutsche Zusammenfassung

Die vorliegende Arbeit beschäftigt sich mit dem Vergleich zweier Testverfahren zur expliziten Messung grammatischer Kompetenz. Es handelt sich dabei um die Grammatikkomponenten des Einstufungstests des Fachbereichs Anglistik/Amerikanistik der Technischen Universität Chemnitz (Technische Universität Chemnitz Placement Test = TUC-PT) und einen Webbasierten Grammatiktest (Web-Based Placement Test = WBPT). Dieser wurde mit dem primären Ziel entwickelt, angehenden Studenten des Magisterhauptfachs Anglistik/Amerikanistik und der entsprechenden Nebenfächer sowie des Kombinationsprofils Fremdsprachen in der Erwachsenenbildung als Übungs- und Vorbereitungsmöglichkeit auf den Grammatikteil des TUC-PT zu dienen.

Der erste Teil der Arbeit gibt einen kurzen Überblick über den Stellenwert computerbasierter Sprachtestmethoden und deren beständig wachsende Bedeutung insbesondere im Bereich der Sprachausbildung an Universitäten und beschäftigt sich mit der Frage der Notwendigkeit solcher Methoden im Bereich Anglistik/Amerikanistik an der Technischen Universität Chemnitz. Aufgrund steigender Studentenzahlen sowie auf Basis des nationalen und internationalen Vergleichs mit der gängigen Testpraxis an anderen Universitäten kann diese Notwendigkeit klar unterstrichen werden.

Ein weiterer Schwerpunkt liegt auf der Erörterung und Diskussion der für die Analyse und den Vergleich der genannten Tests zentralen theoretischen Konzepte. Kapitel 3 versucht daher den theoretischen Rahmen für die folgende detaillierte Behandlung des TUC-PT und des WBPT abzustecken. Besonderes Augenmerk liegt hier auf dem Begriff des Grammatiktestens. Dabei wird in der Hauptsache auf die explizite Leistungsmessung in diesem Bereich im Unterschied zu task-integrierten Verfahren eingegangen. Es konnte herausgefunden werden, dass obwohl explizites und separates Testen von Grammatikkompetenz den Prinzipien des *communicative language testing* in einigen Punkten zu widersprechen scheint, solche Verfahren aber dennoch, in Abhängigkeit vom Verwendungszweck des entsprechenden Tests, durchaus eine Daseinsberechtigung besitzen.

Innerhalb der Diskussion theoretischer Aspekte wird weiterhin ein Schwerpunkt auf die Begriffen *placement testing* sowie *language proficiency* als eines der zentralen Konzepte innerhalb der gesamten Domäne des Sprachtestens gelegte. In diesem Abschnitt wird die Entwicklung der Idee von *language proficiency* von strukturalistisch-behavioristisch definierten bis hin zu kommunikationsorientierten Ansätzen diskutiert. Besondere Bedeutung wird hier dem von Bachman (1990) eingeführten Begriff der *communicative language ability (CLA)* beigemessen, da dieser

Definition von *language proficiency* eine sehr detaillierte und umfassenden Aufschlüsselung der verschiedenen innerhalb des Konzepts kombinierten Kompetenzen zu Grunde liegt. Grammatische Kompetenz als einer der zentralen Interessenschwerpunkte der vorliegenden Arbeit wird dabei klar als eine dieser Kompetenzen verstanden, die jedoch erst in Kombination mit den weiteren von Bachman identifizierten Teilen von *language proficiency* tatsächliche Bedeutung für authentische Kommunikation gewinnt.

Ausgehend von der in Kapitel 3 begonnenen Diskussion des Konzepts *placement testing*, welches generell als einer der möglichen Verwendungszwecke von Sprachtests verstanden wird, beschäftigt sich Kapitel 4 mit der Beschreibung und dem Vergleich des TUC-PT und des WBPT auf Basis eines von Bachman (1990) vorgestellten Klassifikationsschemas, welches aus fünf verschiedenen, jedoch als komplementär zu betrachtenden Kategorien besteht. Diese umfassen 1) den Verwendungszweck (intended use), 2) die inhaltliche Orientierung (content), 3) den Referenzrahmen (frame of reference), 4) das Bewertungsverfahren (scoring procedure) und 5) die Testmethode (test method). Kapitel 4 diskutiert diese Klassifikationskriterien auf theoretischer Basis und wendet sie auf die zu analysierenden Tests an. Dabei wird deutlich, dass beide eine breite Anzahl an Charakteristika gemeinsam haben, sich jedoch insbesondere in den Punkten Verwendungszweck und Testmethode zum Teil starke Unterschiede ergeben.

Ausgehend vom Begriff der Testmethode beschäftigt sich Kapitel 5 mit einer vergleichenden Analyse der Konstruktion von Items in beiden Tests, dabei wird hauptsächlich auf das Testen von Kompetenzen bezüglich der grammatischen Kategorien Zeitformen und Präpositionen eingegangen. Die Beschreibung und der Vergleich erfolgen anhand zweier übergeordneter Kriterien. Diese sind 1) der Aufbau von Items innerhalb verschiedener Varianten der Multiple-Choice Testmethode und 2) die Anwendbarkeit des C-Test-Verfahrens als explizites Messinstrument grammatischer Kompetenz im Vergleich zum traditionellen text-basierten Multiple-Choice Test. Im ersten Teil werden Multiple-Choice Testaufgaben aus dem WBPT, welche auf fortlaufendem, kohärentem sowie meist authentischem Textmaterial beruhen mit zwei der Grammatikteile des TUC-PT verglichen. Ein Hauptaugemerk liegt hier auf dem Einfluss kontextueller Aspekte auf Test Items, die sich mit Zeitformen und Präpositionen beschäftigen. Im zweiten Teil werden C-Testaufgaben aus dem WBPT mit einer text-basierten Multiple-Choice Komponente des TUC-PT verglichen.

Im Hinblick auf das erste der beiden Vergleichkriterien konnte festgestellt werden, dass der direkte Kontext des die besprochen Items umgebenden Satzes in aller

Regel ausreichend Informationen liefert, um die Indenfizierung einer bestimmten Lösung eindeutig zu gestalten. Dennoch fiel auf, dass einem breiteren, möglichst authentischem, Kontext eine sehr bedeutende unterstützende Funktion zukommt. Dies kann vor allem für das Testen von Kompetenz im Bereich Zeitformen bestätigt werden, da hier die Wahl einer bestimmten Konstruktion relativ stark von kontextuellen Faktoren beeinflusst wird. Weiterhin beinhaltet die Bereitstellung eines breiten, kohärenten Basistexts vor allem die Möglichkeit, für ein Testverfahren wie Multiple-Choice, welches in sich bereits bestimmten Einschränkungen, wie z.B. der ausschließlichen Fokusierung auf rezeptive Fähigkeiten, unterworfen ist und welches häufig als unkommunikativ und künstlich im Vergleich mit natürlichem Sprachgebrauch bezeichnet wird, Basismaterialien heranzuziehen, die bestimmte Aspekte natürlicher (geschriebener) Sprache wiederspiegeln. Im Bezug auf satzbasierte Multiple-Choice Items im TUC-PT konnte weiterhin festgestellt werden, dass hier Eindeutigkeit häufig durch Konstruktionen erreicht wird, die einer allgemeinen Vorstellung von natürlichem Sprachgebrauch nicht in allen Punkten entsprechen, wodurch erneut auf die Bedeutung ausreichender kontextueller Einbettung von Test-Items verwiesen werden kann.

Innerhalb des Vergleichs des C-Test Verfahrens im WBPT mit einem Text-basierten Multiple-Choice Teil des TUC-PT ergeben sich einige Nachteile des C-Tests im Vergleich zum traditionellem Multiple-Choice-Testen. Hauptsächlich konnte festgestellt werden, dass für das Testen von Zeitformen und Präpositionen durch C-Test Items eine genaue Definition des Schwerpunkts, wie z.B. das Testen von Kenntnis in der Verwendung einer bestimmten Zeitform im Gegensatz zum Wissen um die formale Bildung einer solchen Konstruktion etc., in vielen Fällen problematisch sein kann. Des weiteren bietet der C-Test im Unterschied zur Multiple-Choice Methode nicht die Möglichkeit, gezielt und kontrolliert auf bestimmte Kontraste und Unterschiede zwischen verschiedenen Zeitformen und Präpositionen einzugehen. Trotz der genannten Probleme fiel innerhalb der Analyse dennoch auf, dass der C-Test im WBPT in Bezug auf die Wiederspiegelung authentischen Sprachgebrauchs durch das verwendete Basismaterial, entscheidende Unterschiede zur analysierten textbasierten Multiple-Choice Aufgabe im TUC-PT aufweist. Für letztere konnte ähnlich wie für die bereits erwähnten satzbasierten Aufgaben eine starke Konstruiertheit festgestellt werden. Besonders für Items, welche ein Hauptaugenmerk auf den Bereich der Zeitformen legen, werden eine sehr hohen Zahl von Hinweiswörtern und Konstruktionen eingebaut, die die Items zwar sehr eindeutig, jedoch auch relativ unnatürlich erscheinen lassen.

Da die Wahl und das Design von Items innerhalb der verschieden Testmethoden

von entscheidender Wichtigkeit für die Validität eines Tests ist, beschäftigt sich die vergleichende Analyse neben übergeordneten Fragen des Einflusses kontextueller Faktoren und der Nützlichkeit des C-Tests als explizites Testinstrument im Bereich englische Grammatik, auch und besonders mit einer detaillierten Beschreibung der einzelnen Zeitformen- und Präpositions-Items. Dabei wird besonderer Wert auf die genauen grammatischen Inhalte, die getestet werden, und, sofern gegeben, die Wahl von Alternativen in Verbindung mit Fragen der Eindeutigkeit der Items diskutiert. Weiterhin werden die für die einzelnen Items erzielten Fehlerquoten dargestellt und, soweit möglich, Probleme auf Seiten der Testpopulation einbezogen.

Neben der detaillierten formalen und inhaltlichen Beschreibung des TUC-PT und des WBPT werden in vorliegender Arbeit auch bestimmte Aspekte, die für eine zukünftige Verwendung des letzteren als Teil des tatsächlichen Einstufungsverfahrens im Bereich der sprachpraktischen Ausbildung innerhalb des Fachbereichs Anglistik/Amerikanistik an der Technischen Universität Chemnitz angesprochen und Möglichkeiten der Verbesserung des Tests diskutiert.

Selbständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Arbeit selbständig und nur unter Zuhilfenahme der angegebenen Hilfsmittel verfasst habe. Alle wörtlich übernommenen Aussagen sind eindeutig gekennzeichnet. Die Herkunft der indirekt übernommenen Formulierungen und Gedankengänge ist angegeben.

Katrin Uhlig

Would you please inform me as soon as possible if EnerCorp is still underlined in doing business with me, or if what has happened here is your way of telling me to do my gasoline business elsewhere.

Sincerely

John Swetnam

The underlined word is used in this paragraph as...

A	○	a gerund.
B	○	a past participle.
C	○	a present participle.
D	○	a present simple passive tense.

http://www.writinghelp-central.com/complaint-letter.html, adapted by K.U.