Univerzita Pardubice

Filozofická fakulta

Katedra anglistiky a amerikanistiky

Testing Reading in English Language Teaching

Diplomová práce

2006   Veronika Benešová

Univerzita Pardubice

Filozofická fakulta

Katedra anglistiky a amerikanistiky

Testing Reading in English Language Teaching

Diplomová práce

2006   Veronika Benešová

Univerzita Pardubice

Filozofická fakulta

Katedra anglistiky a amerikanistiky

Testování čtení ve výuce anglického jazyka

Diplomová práce

Autor: Veronika Benešová

Vedoucí: Mgr. Pavel Brebera

2006

Univerzita Pardubice
Katedra anglistiky a amerikanistiky

Fakulta humanitních studií
Školní rok 2002/2003

ZADÁNÍ DIPLOMOVÉ PRÁCE

pro: VERONIKU BENEŠOVOU
obor: Učitelství anglického jazyka

Název tématu: Testing Reading in English Language Teaching

Zásady pro vypracování:
Diplomantka se ve své práci bude zabývat možnostmi využití jazykových testů řečové dovednosti čtení ve výuce anglického jazyka. Východiskem pro zpracování zvolené problematiky zde bude analýza specifik čtení jako jedné z řečových dovedností společně s diskusí využitelnosti různých typů textů pro žáky jednotlivých úrovní. Následný přehled možných testových úloh zaměřených právě na tuto řečovou dovednost, včetně navrhovaných způsobů hodnocení tohoto typu testů, pak vytvoří potřebný teoretický základ pro část praktickou. V té pak studentka představí výsledky vlastního výzkumného šetření, realizovaného na základní škole, v jehož rámci komplexně zhodnotí soubor testů vlastní konstrukce, především pak efektivitu jednotlivých typů testových úloh a formativní úlohu tohoto typu testů v procesech vyučování/učení anglického jazyka..

Vedoucí diplomové práce: Mgr. Pavel Brebera   Podpis:

Vedoucí katedry: Mgr. Šárka Čáňová, PhD.   Podpis:

Datum zadání práce: 12.5.2003

Tuto práci jsem vypracovala samostatně. Veškeré literární prameny a informace, které jsem v práci využila, jsou uvedeny v seznamu použité literatury.

Byla jsem seznámena s tím, že se na moji práci vztahují práva a povinnosti vyplývající ze zákona č. 121/2000 sb., autorský zákon, zejména se skutečností, že Univerzita Pardubice má právo na uzavření licenční smlouvy o užití této práce jako školního díla podle § 60 odst. 1 autorského zákona, a s tím, že pokud dojde k užití této práce mnou nebo bude poskytnuta licence o užití jinému subjektu, je Univerzita Pardubice oprávněna ode mne požadovat přiměřený příspěvek na úhradu nákladů, které na vytvoření díla vynaložila, a to podle okolností až do jejich skutečné výše.

Souhlasím s prezenčním zpřístupněním své práce v Univerzitní knihovně Univerzity Pardubice.

V Pardubicích 28. 6. 2006

Veronika Benešová

Acknowledgements:

ABSTRACT:

Reading is, together with writing, speaking and listening, one of four essential language skills when acquiring a foreign language. Yet, I feel that many teachers regard its teaching and testing as activity not sufficiently motivating for pupils, not contributing to learning process and therefore not very important.

My thesis draws attention to a very close relationship between teaching and testing in general. Furthermore, it deals with classification of various test types and types of testing. Criteria for a well constructed test (validity, reliability and practicality) are also introduced together with more detailed insight into reading as a language skill. Genres of reading, important skills, sub skills but also strategies employed are also dealt with. As regards constructing tests itself, the readers are familiarised with individual steps in terms of a test's purpose, specifications and choice of test format. The most widely used test formats are dealt with regarding their possible ways of scoring, pros and cons.

The practical part of the paper is based on a research carried out at a primary school. The tests have been constructed on the basis of texts unknown to pupils. I strove to employ such test techniques which the learners may encounter in their future studies, and what is more, which were consistent with the nature of the texts. Subsequently, tests were distributed and completed by pupils. Effectiveness of the used test techniques was measured on the basis of the test results and index of difficulty was chosen as an indicator. Finally, all the test techniques were discussed from view of their suitability.

SOUHRN:

Čtení je spolu s psaním, ústním projevem a poslechem jednou ze čtyř nezbytných jazykových dovedností při studiu cizího jazyka. Nicméně, i přesto cítím, že mnozí učitelé nahlížejí na čtení a také na jeho testování jako na činnost nedostatečně žáky motivující, nijak nepřispívající k ovládnutí jazyka, a tedy i méně podstatnou.

Má diplomová práce upozorňuje na velmi blízký vztah mezi učením a testováním. Dále se zabývá klasifikací různých typů testů a testování. Základní kritéria pro dobře sestavený test (validita, reliabilita a praktičnost) jsou také zahrnuta spolu s hlubším pohledem do oblasti čtení jako jazykové dovednosti. Práce se zabývá také žánry čtení, důležitými mikro a makro dovednostmi, ale i strategiemi, které mohou být

při čtení s porozuměním uplatněny. Co se týče samotného sestrojení testů, čtenáři jsou obeznámeni s jednotlivými kroky, zvláště pak s uvědoměním si cíle testu, jeho specifik a výběru testovací techniky. Jsou probrány nejrozšířenější formáty testů z hlediska výhod a nevýhod tohoto typu testu a jeho možnými způsoby bodování.

Praktická část práce je založena na výzkumu prováděném na základní škole. Testy byly sestaveny na základě textů, které nebyly žákům známé. Snažila jsem se o použití takových testovacích technik, s kterými se žáci v budoucnu mohou setkat, ale také o to, aby testovací technika byla v souladu s charakterem textu. Poté byly testy žáky vyplněny. Také byla posuzována efektivita jednotlivých testových technik na základě indexu obtížnosti, Tento index byl vypočítán podle dosažených výsledků.

Neposledně, všechny použité testovací techniky byly zhodnoceny podle jejich použitelnosti při testování čtení s porozuměním.

TABLE OF CONTENTS

INTRODUCTION

Some teachers see reading as a passive and non-productive language skill, yet we should not neglect it, especially when communicative approach is stressed in the Czech curriculum. The following are samples of negative opinions on reading as introduced in Greenwood:

- I am trying to get through a fairly dense syllabus to equip my students, ultimately, for examinations. I cannot spare the time for frills.
- Reading for pleasure is a private and personal thing. I cannot see how this can be used in the EFL classroom.
- I understand that extensive reading for pleasure can only improve language but I have no way of checking that learning has taken place other than comprehension questions. These activities reduce the pleasure.
- Readers[1] are an expensive luxury. The school cannot afford them. Other things must come first.

(Greenwood 1988:5)

Reading seems to be very unpopular indeed judging after reading the above opinions. Nevertheless, this language skill is an important part of language teaching and learning and I sense that it is not paid as much attention as it indisputably deserves. My assumption lies in the opinion that reading ability does not develop automatically through exposure to the language itself or through practicing grammar or speaking.

From my own experience as a student and as a teacher, I can say that testing can sometimes be rather stressful for both parties. Therefore, I decided to choose the issue of testing reading comprehension to get an insight into this discipline. Moreover, to explain the choice of the reading area, I can support my decision by using Alderson's words, who says:

Reading, through which we can access worlds of ideas and feelings, as well as the knowledge of the ages and visions of the future, is the most enigmatic of the so-called language skills.

(Alderson in Brown 2004:215)

---

[1] Simplified versions of well known literary works. In the Czech Republic used only sporadically due to their high initial outlay and the fact that teachers are not used to working with them.

The object of my thesis is to carry out a research concerning testing reading comprehension with a focus on upper primary school grades. In addition, the paper is aimed at assessment of techniques used for testing reading. The used techniques will be assessed in terms of effectiveness and formative role in teaching/learning processes. Consequently, a set of reading tests was prepared and examined in practice. Finally, the results of the analysis of the testing techniques are described in the practical part of the thesis.

Following on from that, the structure of the paper with its basic outline will be introduced. The theoretical part of my diploma paper is allotted in six chapters. The first three chapters focus on analysing testing in general. Other chapters deal with the aspects of testing reading comprehension in particular.

In the first chapter, I will present very close relation between testing and teaching, reasons for testing in general, basic functions of tests and what positive and negative effects of testing we may encounter.

The second chapter analyses types of tests, differentiated according to their purpose and types of testing. In other words, proficiency, achievement, aptitude, diagnostic, and placement tests together with direct, indirect, discrete point, integrative point, norm-referenced, criterion-referenced, subjective and objective testing will be discussed.

The next chapter focuses on the most essential considerations in test design. It will give suggestions regarding construction of reliable, valid and efficient tests.

The fourth chapter provides readers of my thesis with reasons for employing and testing reading, it also presents genres of reading together with its micro and macro-skills and reading strategies. Furthermore, this chapter identifies the four basic types or class reading performances – reading aloud, silent reading, intensive and extensive reading.

The following chapter of the theoretical part describes stages of a test construction in particular with suggestions regarding design of good reading tests as well as advice for a good test development. Furthermore, this chapter refers to the steps of composing test specifications.

Eventually, the last chapter deals with presenting different types of reading test formats. When discussing the formats, their advantages, disadvantages and approaches to their scoring will be taken in account as well.

To conclude with the structure of the practical part of my paper, it is organised in connection with the individual test analyses. Background of the research, concerning choice of the texts, aspects of test administration and other issues is introduced at the beginning. Next, analysis of the six tests in order to assess the used testing techniques is provided in the conclusion. Since one part of the test analysis contains estimation of difficulty by means of measuring difficulty index, the method is described in detail. Samples of completed tests, their scores, grades and results of the measurement are submitted in the appendix.

1. TESTING, TEACHING AND FUNCTIONS OF TESTS

1.1. TESTING AND ITS RELATION TO TEACHING

Heaton suggests a need to separate testing from teaching. Nevertheless he declares that "both testing and teaching are so closely interrelated that it is virtually impossible to work in either field without being constantly concerned with the other" (Heaton 1988:5).

Testing plays a very significant role in the evaluation process. I feel that it is necessary to make clear distinction among the following three terms which are very closely related – evaluation, assessment and testing.

Slavík analyses these terms in Hodnocení v současné škole and he points out that the use of the above mentioned terms in Czech may be rather confusing. Therefore, under the influence of English, it is necessary to distinguish what the difference between evaluation and assessment is. Duncan and Dunn consider evaluation to be a result of assessment – marking, awarding a diploma, or a written expert's report (Duncan in Dunn in Slavík 1999:35). Moreover, Harris declares that evaluation "involves looking at all the factors that influence the learning process, such as syllabus objectives, course design, materials, methodology, teacher performance and assessment" (Harris et al. 1994:2).

On the other hand, when assessing, teachers must measure the performance of the learners and the progress they make. Pasch sees assessment as a process of gathering information serving as basis for evaluation – testing, examining or diagnostic observation of pupils (Pasch in Slavík 1999:35).

According to Slavík, assessment has three main functions. They are: motivational (closely related to emotional side of assessment), cognitive (applies to intellectual or rational assessment) and conative (performs a human will to act) (Slavík 1999:16-18).

In addition, assessment may be both holistic and analytic. Holistic assessment is spontaneous, unintentional and emotional. It is based on overall impression of the object and can be either ´good´ or ´bad´. Analytic assessment is intentional, consciously focused and may be analysed and formalised. Results of formalised assessment are marks (Slavík 1999:34-35).

Furthermore, Slavík distinguishes between summative and formative assessment. Aim of the first is to get a final overview of the performance results and the fundamental purpose is to determine if "passes – does not pass" (Slavík 1999:37). Typical examples of summative assessment are university entrance exams or mid-term school exams. However, formative (corrective) assessment provides feedback after a certain performance when there is a high chance of even enhancing the performance in future. In other words, assessment of such type helps teachers and learners to explore better ways to reach the aim. (Pasch in Slavík 1999:38). Formative assessment is considered helpful, it is an instrument for achieving better results and self-recognition. Subsequently, it positively affects pupils´ motivation (Slavík 1999:39). Nevertheless, we must not see teachers who carry out formative assessment as teachers who work at ´schools without tests and grades´. Even testing and grading may be considered formative assessment as long as the tests and their results are informative and corrective (pupils learn how to monitor their work, they start to take responsibility over their studies, they take pleasure in their success etc.) (Petty in Slavík 1999:40).

Harris, on the other hand, differentiates formal and informal assessment. Formal assessment is, according to him, identical to testing and establishes test or exam conditions, as opposed to informal assessment which is a way of collecting information about learners´ performances in normal classroom conditions (Harris et al. 1994:2).

1.2 BACKWASH

As regards the relationship between teaching and testing, Hughes mentions that we should not understand testing to be a servant of teaching but there should rather be a partnership between the two (Hughes 1989:44-46).

According to Hughes, backwash is "the effect that testing has on teaching and learning" (Hughes 1989:1). There are two types of backwash and it is either beneficial (positive) or harmful (negative).

Hughes states that negative effects are likely to appear when content and formats of a test differ from the objectives of a course. In other words, when teachers produce tests testing language skills, sub skills[2] and subject matter that had not been covered during the course, they can under no condition achieve beneficial feedback (ibid.).

---

[2] Some ELT experts use terms ´micro/macro skills´ (Brown). See chapter 4. 3

Alderson shares the same opinion when he states: "If we do not test what we are trying to teach, there is a danger that students will not pay much attention to what is taught, or that teachers will cease to teach it. And we certainly do not want to undermine what teachers are teaching" (Alderson in Nuttall 2005:220).

On the contrary, by teaching and learning precisely the ability that teachers want to encourage, positive feedback will be provided. Furthermore, Hughes adds the following. "There is a tendency to test what it is easiest to test rather than what it is most important to test" (Hughes 1989:44). He supports his opinion by claiming that many teachers often prepare grammar tests which are objectively (thus relatively easily) scored rather than creating tests which are scored subjectively (Hughes 1989:116). In order to achieve positive feedback, he suggests preparing criterion-referenced tests[3]. He thinks that "if learners perform the task at the criteria level, then they will be successful on the test, regardless of how other learners perform" (Hughes 1989:45).

The last, however not the least criteria of achieving beneficial backwash is to make tests as comprehensible as possible in terms of rationale, specifications and sample items, especially when there is a new testing format introduced. Learners should always be aware of what exactly the test demands of them (Hughes 1989:46).

1.3 FUNCTIONS OF TESTS

According to Penny Ur, there are four main functions of tests. Firstly, tests are supposed to provide teachers with information about where the students are at the moment so that they can decide better about what to teach next. Secondly, tests should give students information about how much they know so that they can be aware of what subject matter they need to learn or review. Thirdly, tests may serve as a tool reinforcing a noisy class to keep quiet and to concentrate. Another crucial function of a test is that they provide students with a sense of achievement and progress in their learning (Ur 1996:36).

Another view on functions of language tests is expressed by Madsen, who claims that English tests are beneficial to mastering the language. In particular, tests discover what each student has learned, and they point up which language items should be studied further. Consequently, by emphasising the course objectives, the pupils can

---

[3] See Chapter 2 - Types of testing

set their personal goals. Yet another important function of a test is to locate the precise areas of student's weaknesses and strengths (Madsen 1983:4).

Furthermore, Heaton mentions that a test should be concerned with evaluation for the purpose of enabling teachers to increase their own effectiveness by making adjustments in their teaching to enable certain groups of learners or individuals in the class to benefit more (Heaton 1988:6). The same opinion is shared by Madsen who declares that teachers can use tests to diagnose not only efforts of students but also those of their own (Madsen 1983:5).

In addition, Heaton claims that a well made test of English can help in motivating learners (Heaton 1988:7). However, only such tests which are of appropriate difficulty, are announced in advance and cover the skills issued in the course may lead to creating positive intrinsic motivation (Madsen 1983:4).

A very important function of a test is evaluation of students' performance. Nevertheless, Heaton mentions that evaluation for purposes of comparison or selection is only one of the reasons for testing and should not be the sole purpose of testing in schools. In other words, testing is only one part of evaluation of learners but has a very important role since it provides feedback about the results of education for teachers as well as for learners (Heaton 1988:6)

2. TYPES OF TESTS AND TESTING

In language testing various types of tests are designed due to the fact that learners are assessed for various purposes. Therefore, it is crucial to know exactly what the aims of a test are. Although confusion regarding terminology used to design these types of tests still exists among test-makers, this chapter will try to clarify their use. In addition, different types of testing will be dealt with.

2.1 ACHIEVEMENT TESTS

According to Madsen, achievement tests measure mastery of language sub skills or mastery of the skills which have been presented in the entire course (Madsen 1983:9). In other words, it means that these tests measure whether the objectives of a course have been successfully achieved.

Heaton subdivides this group of tests into class progress and final achievement tests. Class progress tests intend to "measure the extent to which the students have mastered the material taught in the classroom" and are, according to him, the most widely used. They are based on the evaluation of the class teacher, which depends on the teacher's knowledge of his or her students, the programme which has been studied, and the teacher's own goals and aims of teaching (Heaton 1988:171).

Similarly to Heaton, Hughes also distinguishes final achievement tests from class progress tests. He declares, that final achievement tests are administrated at the very end of a course and must correspond with the course syllabus, with the textbook, and other materials used throughout the course. This procedure is referred to as the ´syllabus-content approach´ (Hughes 1989:11).

Nevertheless since even ´syllabus-content approach´ tests may be badly designed, Hughes recommends using an alternative approach. ´Course-content-based´ tests are according to him more accurate in terms of information about individual and group achievement and they are more likely to produce more beneficial washback (ibid.).

2.2 PROFICIENCY TESTS

Proficiency test, according to Heaton, are designed in order to define students proficiency in a certain language with reference to a particular task which the student will be asked to perform in the future (Heaton 1988:172).

The aim of proficiency tests is to design such tests which will be measuring specific language skills and sub skills that will be demanded on the student by their future job or by next course of study (Heaton 1988:173). In other words, proficiency tests are designed to show what the proficiency level of their candidates is.

However, Hughes indicates that there may emerge hesitations as to what the expression ´to be proficient´ in language really means. According to him, there are two types of proficiency tests. Firstly, there are tests designed to discover whether a candidate has "sufficient command of the language for a particular purpose". An example of this test would be an entrance test for foreign students to follow a university course in Great Britain. Secondly, there are proficiency tests whose function is to show into what extent a candidate has reached a certain standard. However, these tests do not bear in mind any purpose for the language and they only indicate how the successful candidates have mastered the language (Hughes 1989:9-10).

2.3 DIAGNOSTIC TESTS

Diagnostic test are, according to Hughes, used "to identify students´ strengths and weaknesses". Their task is to disclose and diagnose areas of difficulties so that appropriate actions for improving this situation can be later taken (Hughes 1988:173). When constructing a diagnostic test a vast number of items have to be included in order to get detailed information about the student's knowledge of a particular language skill. Hand in hand with a large number of examples in the test goes impracticality of administrating these tests and that is why, as Hughes suggests, only few tests are composed for solely diagnostic purposes (Hughes 1989:13-14).

2.4 PLACEMENT TESTS

"Placement tests are intended to provide information which will help to place students at the stage of the teaching programme most appropriate to their abilities"

(Hughes 1989:14). In other words, their task is to assign students of languages to different classes at different levels and thus save time and effort of teachers.

2.5 APTITUDE TESTS

John Heaton mentions yet another type of test which stays rather unnoticed by other test experts and that is aptitude test. Function of such tests is "to assess aptitude for learning a language" (Heaton 1988:173). That is to say, they are constructed to measure the candidates' probable performance in a language which they have not started to learn yet. Since aptitude for languages consists of many factors, for example age, intelligence, motivation, memory, sensitivity to grammatical and phonological patterning, it is impossible to take all these factors in account when constructing this type of test. One way to a better designing of aptitude tests is, according to Heaton, focusing on various aspects according to the specific tasks for which a candidate is being trained (ibid.).

2.6 DIRECT VERSUS INDIRECT TESTING

Hughes says that it is necessary to distinguish between these two approaches to test construction. According to him, direct testing requires "the candidate to perform precisely the skill which we wish to measure" (Hughes 1989:15). The crucial thing for test writers is to make the tasks, texts, listening exercises and so on, as authentic and thus as realistic as possible (ibid.).

On the other hand, Hughes claims that indirect testing only aims to measure the abilities which underline the skill we are interested in. Moreover, he sees indirect testing superior to direct testing because its results are more general. He supports his opinion by claiming that indirect testing provides a representative sample of a particular structure as opposed to direct testing which is limited to a rather small sample of tasks and therefore does not include a representative structure (Hughes 1989:16). Direct testing has many attractions, even though there may sometimes appear problems connected with reliability[4] of testing productive skills. However, if the test writers are "clear about just what abilities they want to assess, it is relatively straightforward to

---

[4] See Chapter 3.2 - Reliability

create the conditions which will elicit the behaviour on which to base their judgement"
(Hughes 1989:15).

2.7 DISCRETE POINT VERSUS INTEGRATIVE TESTING

According to H. Douglas Brown:

> Discrete point tests are constructed on the assumption that language can be broken into its component parts and those parts can be successfully tested. These components are basically the skills of listening, speaking, reading, and writing, and various hierarchical units of language (discrete points of phonology/graphology, morphology, lexicon, syntax, and discourse.
>
> <div align="right">(Brown 2004:8)</div>

Only those proficiency tests which included samples of all above mentioned skills and as many linguistic discrete points as possible, were considered efficient. However, with the era of emphasising communication, authenticity and context, many ELT methodologists criticised this opinion. One of them was John Oller, who supports integration (hence "integrative" testing) by saying that language competence is a unified set of interacting abilities and therefore must not be tested separately (Oller in Brown 2004:8). In other words, integrative testing makes the test-taker combine many language elements for successful completion of a task.

To conclude, according to Harris, both integrative and discrete point testing have their advantages and disadvantages. In order to create a valid[5], reliable, and practical[6] test with beneficial washback, the test-makers should combine both formats that is discrete item tasks for testing receptive skills and integrative tasks especially for productive skills (Harris et al. 1994:35).

2.8 NORM-REFERENCED VERSUS CRITERION-REFERENCED TESTING

We use a norm-referenced test to show how a student's performance compares with the performance of the other students in the same group (Heaton 1988:16). The purpose of such tests is to place test-takers along a mathematical continuum in rank order. Scores of norm-referenced tests are reported back to the learners in the form of a numerical score (for example, 100 out of 200) or a percentile rank (such as 50 percent, which means that the test-taker's score was better than 50 percent of the total number of

---

[5] For validity see chapter 3.1
[6] For practicality see chapter 3.4

test-takers, but worse than 50 percent in that administration) (Brown 2004:7). However, according to Hughes, we are not directly told what the student is able to do in the language (Hughes 1989:17).

On the other hand, "we use criterion-referenced tests to find out whether a student can perform a particular task or not" (Heaton 1988:18). Learners are classified according to whether or not they have performed the task satisfactorily by ´pass´ or ´fail´. According to Hughes, this means that "students are encouraged to measure their progress in relation to meaningful criteria, with feeling that, because they are less able than most of their fellows, they are destined to fail" (Hughes 1989:18). In other words, criterion-referenced tests set standards meaningful in terms of what learners can do and they motivate them to achieve those standards. Finally, since criterion-referenced tests are more likely to provide learners with beneficial washback effect, they are often preferred to norm-referenced tests (ibid.).

As regards design, construction, and development of these tests, Bachman points out that there are crucial differences. As for criterion-referenced tests, these are designed to represent a specific level of ability in a foreign language, and the items of the test are selected adequately to the level. On the other hand, norm-referenced tests are designed to maximise distinctions among test-takers, which means that the items are selected according to how well they distinguish individuals performing poorly or well on the test (Bachman 1990:75-76).

2.9 OBJECTIVE VERSUS SUBJECTIVE TESTING

Regarding scoring tests, there are two methods. Whereas objective testing requires no judgement by the examiner (because an objective item has only one or a limited number of correct answers, and such tests can be marked by an inexperienced teacher or marker, even by a machine), scoring subjectively requires teacher's judgement when awarding marks (since subjective questions allow test-takers much more freedom and flexibility in the answers, they can only be marked by a competent teacher or marker) (Heaton 1988:31-32). In other words, it means that no matter which teacher marks the test, a candidate will score the same mark, under the condition that they are tested objectively.

Nevertheless, according to Heaton, objective tests require plenty of preparation time and he proves it by saying that when creating an objective test "the tester spends a great deal of time constructing each item as carefully as possible, attempting to anticipate various reactions of the test-takers at each stage" (Heaton 1988:26).

As regarding high reliability of scoring tests, ELT experts say that it is mostly achieved by scoring objectively. However, there are exceptions and an objective test will not be reliable if:

- The test items are poorly written
- Irrelevant areas and skills are emphasised in the test simply because they are ´testable´
- It is confined to language-based usage and neglects the communicative skills involved.

(Heaton 1988:27)

Even though objective tests are often criticised for being easier to answer than subjective tests, items in an objective tests may be difficult to answer as well. However, the following is what Heaton declares: "The fact that objective tests may generally look easier is no indication that they are easier" (Heaton 1988:26). On the other hand, in some cases, it is better to score subjectively since subjective questions allow greater freedom and flexibility.

To sum up, it is useful to find certain balance between these two scoring methods because, as Heaton states that "most good tests contain both objective and subjective types of items" (Heaton 1990:33).

3. THE MAIN TEST CRITERIA

The three main criteria on a well prepared test, validity, reliability, and practicality will be described in this chapter.

3.1 VALIDITY

In Heaton's opinion "validity of a test is the extent to which it measures what it is supposed to measure and nothing else" (Heaton 1988:159). This chapter will present various types of validity. When discussing face, content, and construct validity, testing reading will be especially taken in account. Furthermore, other types of validity such as washback and criterion-related will be also dealt with.

As Hughes suggests, a test is to said to have face validity "if it looks as if it measures what it is supposed to measure" (Hughes 1989:27). Teachers should always let other colleagues or friends examine their test, since what may seem right for one person does not necessarily seem right for another person. Only by careful viewing of the test by someone "impartial" can some ambiguities and absurdities of a test be discovered and eliminated (Heaton 1988:159). Anastasi points out that "face validity pertains to whether the test ´looks valid´ to the examinees who take it" (Anastasi in Weir 1990:26). Face validity is similarly viewed by Heaton who claims that if learners do not accept their test as valid, their reaction may be that the do not put maximum effort into performing the tasks (Heaton 1988:160).

The test has content validity if its content, according to Hughes, consists of a representative sample of the language items, which it is meant to test (Hughes 1989:22). In order for the test to have this type of validity, according to Heaton, it is crucial to propose a chart of specifications of the test, which would precisely describe the particular language skills and areas to appear in the test (Heaton 1988:160-161).

Similarly Anastasi provides a set of suggestions for establishing content validity when she claims the following: "The behaviour domain to be tested must be systematically analysed to make certain that all major aspects are covered by the test items and in the correct proportions" (Anastasi in Weir 1990:25).

Construct validity is, according to Cronbach, the most important one and he supports his argument by saying that every time an educator asks "but what does the

instrument really measure?" he or she is asking for information on construct validity. In other words, he or she wants to know if the test measures exactly the ability and skills, which it is supposed to measure (Cronbach in Weir 1990:22).

Morrow declares that the difficulties of determining exactly what it is that needs to be measured have recently lead to a new term – washback validity (Morrow in Weir 1990:27). That is to say, this type of validity illustrates "the washback of the test on teaching and learning that precedes it" (Porter in Weir 1990:27).

To summarise the types of validity described in the literature concerned, Weir mentions the following: "For construct, content, face and washback validity knowing what the test is measuring is crucial" (Weir 1990:27).

However, there is another type of validity where knowing precisely what the test measures is not important to such extent – criterion-related validity. It compares test results with results provided by highly dependable and independent assessment of learner's ability. According to Hughes, there are two types of criterion-related validity - concurrent and predictive. Firstly, concurrent validity may be achieved when the test and the criterion are administered at around the same time. In that case, validity is proved when the comparison between the two sets of scores reveals a high level of agreement. The higher level of agreement there is, the higher concurrent validity there is. Secondly, predictive validity concerns the degree to which a certain test can predict test-taker's future performance. A proficiency test predicting a foreign student's ability to cope with a graduate course at a British university would be an example of a test showing predictive validity (Hughes 1989:23-25).

3.2 RELIABILITY

According to Madsen, a reliable test is "the one that produces essentially the same results consistently on different occasions when the conditions of the test remain the same" (Madsen 1983:179). However, Hughes points out that achieving a hundred percent reliability of tests is not possible due to several factors, and he supports his argument by presenting an example of two tests taken on two different days. Even if all the circumstances of administrating the tests would be identical, if the scoring of the test called for no judgement, if no learning or forgetting had taken place during these two days, it would still be impossible to achieve the very same results. He thinks that this is

due to the fact that people never behave in exactly the same way on every occasion. By no means, this means that test scores should not be trusted (Hughes 1989:29).

Weir adds another two aspects of reliability. Firstly, the learner's score should remain the same even if there are different evaluators of the particular test. Similarly, the score should stay the same abstractedly from which version of the test the learner takes (Weir 1990:31). As Heaton suggests, there are many factors affecting reliability of a test:

- The extent of the sample of material selected for testing. The larger the sample, the greater the probability that the test as a whole is reliable
- The administration of the test: is the test administrated to different groups under different conditions or at different times?
- Test instructions: are the various tasks expected from the test-takers made   clear to all candidates in the rubrics?
- Personal factors such as motivation and illness
- Scoring the test: subjectively scored tests sometimes face problems with reliability as opposed to objectively scored tests.

(Heaton 1988:162-163)

Hughes provides suggestions how to achieve test reliability as high as possible. Firstly, similarly to Heaton, Hughes points out the more items appear in a test, the more reliable that test will be. A very crucial thing to remember, however, is that supplementary items should be independent of each other as it prevents learners from answering the second question correctly if they got the original question wrong. He mentions that this problem frequently appears in many reading comprehension tests, where there are additional questions. Secondly, learners should not be granted too much freedom in choice of questions and in the way they answer the ones that they have selected. Thirdly, test writers should avoid presenting ambiguous items whose meaning is not clear or to which exists another acceptable answer. Another advice from Hughes is that, when creating a test, test-makers must always try to present clear and explicit instructions. Unless the instructions are one hundred percent exact, the test writers can be sure that there will always be at least one student who will misunderstand what he/she is asked to do (Hughes 1989:37-40).

Another factor influencing reliability is familiarity of the candidates with format and testing techniques. Finally, layout and legibility of written tests are also factors

influencing reliability. Hand in hand with these factors goes the need of providing non-disturbing and uniform conditions of the test administration (ibid.).

As far as the factors of score reliability are concerned, Hughes suggests to use such items that permit scoring which is as objective as possible, make comparisons between candidates as direct as possible, provide a detailed scoring key, train scorers, and identify test-takers by numbers not names (Hughes 1989:40-42).

In addition to measuring test reliability, Heaton mentions three criteria. Firstly, administer the very same test after some period of time again. Comparison of the results will consequently show into what extent the test is really reliable. Secondly, teachers can ensure reliability by administrating two similar versions of a particular test in terms of sampling, difficulty, and length to the same group. The more similar the results are, the higher reliability of the tests is indicated. The third means is termed split-half method. It is based on splitting the test items into two identical parts. The results of one part should correlate with the results of the other part and thus acknowledge reliability (Heaton 1988:162-163).

As for communicative language testing, Heaton suggests that there has recently been introduced another dimension to the concept of reliability known as profile reporting. He states that: "In order to obtain a full profile of a student's ability in the target language, it is necessary to assess his or her performance separately for each of the different areas of communication", which in other words means that students need to be examined in all language skills (Heaton 1988:163).

3.3 VALIDITY VERSUS RELIABILITY

As stated above, validity and reliability are the major criteria for evaluation of tests. All test design experts agree that tests should be both valid and reliable. However, according to Heaton, there is a conflict between these two criteria and he supports his opinion by declaring that: "The greater the reliability of a test, the less validity it usually has" (Heaton 1988:164). Therefore, test makers have to decide whether they will increase validity at the expense of reliability or vice versa. This problem causes, according to Guilford, tension between validity and reliability because while a test can be reliable without being valid, it is impossible to have a valid test which is not reliable (Guilford in Weir 1990:33). In other words, a test can consistently indicate the same

results even though it does not measure what it should measure. By contraries, if a test does not consistently measure something, it is inevitable that the accuracy of such measurement is low (Alderson 1995:187).

To sum up the opinions on the tension between validity and reliability, Bachman states the following. "Neither is absolute, in that we can never attain perfectly error-free measures in actual practice" (Bachman 1990:26).

3.4 PRACTICALITY

Hughes states that even "the best test may give unreliable and invalid results if it is not well administrated" (Hughes 1989:152). According to Weir, there are several factors which can turn a valid and reliable test into a poor one due to its inefficiency in terms of economy, administration, scoring and interpretations of results (Weir 1990:34). Hughes describes other factors such as decent materials and equipment, clear instructions for the test-takers as well as for the proctors and invigilators, and appropriate rooms in which tests are held (Hughes 1989:152-154).

Furthermore, every test should be economical in terms of cost and time. These two factors are closely related as Weir demonstrates. "The longer it takes to construct, administer and score a test, and the more skilled personnel and equipment that are involved, the higher the costs are likely to be" (Weir 1990:34).

4. TESTING READING

> Teachers often feel that they do not know how to test reading. However, the difference between testing and teaching is not as great as people believe. If there is a difference, it is not so much in the materials themselves but in the way they are used and the purpose for which they are used.
>
> (Alderson in Nuttall 2005:212)

4.1 REASONS FOR READING AND ITS TESTING

The primary purpose why we read (besides reading for language learning) is because: "we want to get something from the writing". The message we want to get might be facts, ideas, feelings etc. Transfer of such messages is considered ´an authentic reason´. In other words, we are not concerned with language learning but with the uses of reading in our daily lives outside the classroom (Nuttall 2005:3).

When looking at reading with regards to language learning, it is generally believed that reading in mother tongue helps to build vocabulary, contributes to expressing oneself and therefore it encourages reading especially in the young age. The same applies to learning a foreign language. From the methodological point of view, reading reinforces and extends vocabulary because, according to Brown, there is much larger variety of words in written than in spoken language (Brown 1994:290).

Reading "is a skill that is taken for granted by many teachers" because in literate societies in the world, majority of people learn to read by the age of six or seven. When teaching a foreign language, teachers automatically expect their learners to acquire this skill as well. Even young-learner's textbooks in a foreign language presuppose a learner's ability to understand written texts only because it is a book that is the medium of the transfer process. On the contrary, reading in a foreign language comes by no means so naturally and therefore learners should not be exposed to written texts without any instructions and systematic training of this language skill (Brown 1994:290).

Many ELT methodologists consider reading to be the most long lasting and the most essential of the four basic language skills concerning success in all educational contexts. Brown supports his opinion by claiming that written word is often used as a

stimulus for test-takers responses and he adds that even some oral interviews can involve reading for particular tasks (Brown 2004:185).

As ELT experts suggest, learners must encompass two primary hurdles in order to become efficient readers. Firstly, there is the hurdle of mastering fundamental bottom-up strategies for processing separate letters, words, and phrases as well as mastering of top-down strategies, important for comprehension of texts. Secondly, there is a need for development of appropriate content and formal schemata in order to accomplish the task successfully. In other words, students must acquire adequate background knowledge and cultural experience (Brown 1994:284).

According to Kitao, when reading a text, the more background knowledge learners have, the easier, more complete, and more accurate the comprehension is. However, the text should never be about a subject that the test-takers are very familiar with since it would be too easy to answer the questions without even comprehending the test fully (Kitao&Kitao).

Comprehension of a certain text is not the sole aim of assessing reading, as Brown suggests. Various strategic pathways which learners use in order to achieve ultimate understanding also play a very important role. The decision which test-takers´ strategies need to be enhanced is therefore solely in hands of teachers (Brown 2004:185).

4.2 GENRES OF READING

According to Brown, it is necessary to mention different types of genres of written texts since "it is the genres which ultimately form part of the specifications for assessments of reading ability" (Brown 2004:196). He divides genres of reading into three groups:

1. Academic reading
   - general interest articles (in magazines, newspapers)
   - technical reports (lab reports, professional journal articles)
   - reference material (dictionaries)
   - textbooks, theses, essays, papers, test directions
2. Job-related reading
   - messages (phone messages)
   - letters/emails
   - manuals, directions, directories, financial documents
   - reports (job evaluations, project reports)

3. Personal reading
- newspaper and magazines
- letters, emails, greeting cards, invitations, messages, notes, lists
- schedules, recipes, menus, maps, calendars, advertisements
- novels, short stories, jokes, drama, poetry
- questionnaires, forms, medical reports, immigration documents

(Brown 2004:186-187)

These genres all have their own set of governing rules and conventions for their manifestation and readers must be able to anticipate these conventions in order to know what exactly to focus on within the text and consequently process the meaning correctly (Brown 1994:286).

What Brown calls ´anticipation´ of conventions of various genres, Nuttall calls ´prediction´ of structures of various genres. People tend to think in similar ways and therefore we, as readers, are able to predict (anticipate) what to expect from a text. The ability to predict (anticipate) exists because we, as readers, understand the text. However, it is very difficult to determine into what extent the ability to use prediction (anticipation) in order to understand exists (Nuttall 20005:118).

4.3 READING MICRO AND MACRO-SKILLS AND READING STRATEGIES

When assessing reading comprehension, it is also important to take the following reading micro and macro-skills in consideration. Brown sees micro-skills inferior to macro-skills and he mentions that micro-skills serve as a means of improving macro-skills (Brown 2004:187).

According to Brown, the most important micro-skills involved in reading are:

- Discriminating among the distinctive graphemes and orthographic patterns of English.
- Retaining chunks of language of different lengths in short-term memory
- Recognising a core of words, and interpreting word order pattern and their significance.
- Recognising grammatical word classes, systems, patterns, rules.
- Recognising that a particular meaning may be expressed in different grammatical forms.

(Brown 2004:187-188)

On the other hand, the most important macro-skills are:

- Recognising the rhetorical forms of written discourse
  and their significance for interpretation.

- Recognising the communicative functions of written texts, according to form and purpose.
- Inferring links and connections between events, deducing causes and effects, detecting such relations as main idea, supporting idea, new information, given information, generalisation, example, comparison, contrast, opinion, and support anticipation what will come next.
- Distinguishing between literal and implied meanings.
- Deducing the meaning and use of unfamiliar lexical items through contextual clues.
- Understanding relationship within the sentence, signalled by lexical devices, anaphoric reference, connectives linking sentences and ideas.
- Varying reading strategies according to the type of text and the purpose of reading.

(Brown 2004:187-188)

When Brown points out development and use of different reading strategies leading to efficient comprehension he mentions skimming and scanning in particular. He justifies his opinion by the fact that he sees these two strategies the most valuable for learners (Brown 1994:293).

The same opinion is shared by Duffy and Roehler who state:

> Reading rate is one of the keys to efficient study. Therefore, instead of reading every single word, skilled readers scan to find potentially helpful information and subsequently skim to locate key words that signal the information the need"

(Duffy and Roehler 2004:183)

Skimming is, according to Nuttall "process of rapid glancing through a text either to search for a specific piece of information (a name, a date) or to get an initial impression of whether the text is suitable for a given purpose" (Nuttall 2005:49). In other words, skimming is a strategy that gives readers sense of the theme, purpose, and organisation of the text, and perspective or point of view of the writer. After skimming a text, learners should be able to answer the following questions:

- What is the main idea of this text?
- What is the author's purpose in writing the text?
- What kind of writing is this (newspaper article, novel, manual)?
- What type of writing is this (narrative, technical)?
- How easy or difficult do you think this text will be?
- What do you think you will learn from the text?
- How useful will the text be for your profession (needs, interests)?

On the other hand, scanning is "the process of rapid glancing through a text to determine its gist or to keep ourselves superficially informed about matters that are not of great importance to us" (Nuttall 2005:49). In other words, mastering of scanning strategy requires rapid identification or relevant pieces of information from following types of texts: "a one- to two-page news article, an essay, a chapter in a textbook, a technical report, a table of chart depicting some research findings, a menu, and an application form" (Brown 2004:209).

There is an endless number of scanning objectives, among which belong locating data such as: "a date, name, or place, results reported in a specified cell in a table, specified data needed to fill out an application, the principal divisions of a chapter" (Brown 2004:209).

When testing scanning, Simmons and Palmer suggest first giving learners the reading task, and then presenting them with the text. They see this sequence as very crucial since it helps the learners in a sense that they know what information exactly they are searching for before they try the deal with the text itself (Simmons and Palmer 1988:32).

With regards to testing scanning, one of the purposes of scanning is to identify relevant information as rapidly as possible. Therefore test-markers may sometimes decide to calculate timing into their scoring procedures (Brown 2004:209).

4.4 TYPES OF CLASS READING PERFORMANCE

As well as other language skills, reading can also be subdivided into different types of performances. As regards reading, variety of performance is based on multiplicity of text genres rather than on the variety of apparent types of reading performance. However, when speaking about assessment procedures, the following types of reading performance need to be specifically identified. These serve test-makers as means for setting reading tasks (Brown 2004:189).The following graph shows types of class reading performance as seen by Brown:

Classroom reading performance

Oral   Silent

Intensive   Extensive

Linguistic   Content   Skimming   Scanning   Global
(Brown 1994:297)

4.4.1 READING ALOUD AND SILENT READING

Reading aloud in classrooms is not very popular with many ELT methodologists. For example Nuttall claims that "reading should be practised mainly for comprehension and not for the sake of pronunciation". In other words, when reading aloud, learners must concentrate on correct pronunciation and therefore it is difficult for them to focus on recognising the meaning as well (Nuttall 2005:58).

Similarly, Ytreberg and Scott point out problems connected with reading aloud. Firstly, texts read aloud are of little interest for the rest of the class. Secondly, they may encourage stumbling, stammering or making mistakes in pronunciation. Thirdly, such texts are demanding in terms of time and to conclude, they discourage learners from practising silent reading techniques (Ytreberg and Scott 1991:57).

Silent reading is preferred to reading aloud in most methodological books concerning reading comprehension. The reason for it is that when reading silently, learners need not think about correct pronunciation and therefore they can absorb more information about the text. What is more, readers are allowed to read at speed which is convenient for them, they can return to parts which they did not understand, and they can correlate facts learnt within the text.

4.4.2 INTENSIVE AND EXTENSIVE READING

Silent reading may be, according to Brown, subcategorised into intensive and extensive.

By intensive reading, he understands deep and thorough study of a text. Intensive reading demands thorough understanding of linguistic comprehension (grammatical, symbolic and lexical forms) as well as mastering of content-related

reading. Sometimes, meaning can be "trapped" inside the text and need not necessarily be cognisable at first glance. However, a skilled reader should be able to extract meaning from the text (Brown 1994:297).

On the other hand, extensive reading is a term for more general comprehension of a certain text. The term applies to texts longer than one page and includes pleasure, technical, scientific and professional reading. This type of reading performance may, at times, involve skimming and scanning strategies which, together with pre-reading activity, serve as means for gaining global or general meaning of the text (Brown 1994:298).

Nevertheless, Greenwood points out that mastering of intensive and extensive reading are by no means skills which can develop overnight. On the other hand, effective use of these two types of reading are long time tasks which demands alternating of both (Greenwood 1988:9).

5. STAGES OF READING TEST CONSTRUCTION

It is needles to say that reading skills are often tested together with the other three language skills. Heaton is of the opinion that reading is in many tests most closely related to writing and he presents the following arguments to prove it.

Firstly, reading provides stimuli for writing (an example may be required to write a reply to a letter. In order to respond to this letter, learners must be able to understand the letter after reading it).

Secondly, reading can in many cases be integrated into the writing skill itself (in order to complete an application form, students must understand the written data from the form).

Thirdly, students may be asked to read a text in order to take notes. However, when testing project work, which is mostly based on certain amount of initial reading, students´ reading skills can be integrated with writing, speaking and possibly listening skills as well (Heaton 1990:92).

According to Brown, most classroom teachers of English come in contact with placement, diagnostic or achievement tests during their career. And it is exactly the necessity to determine the right type of test that he sees as very crucial in order to help teachers focus on the specific objectives of the reading test (Brown 2004:43).

However, there are other practical steps which every teacher should follow if he or she wants to create and administer interactive and intrinsically motivating tests. As Brown suggests, the biggest problem of many teachers is the fact that they try to think of "a wonderfully innovative instrument that will garner the accolades of their colleagues and the admiration of their students" (Brown 1994:386) and that causes more damage than good. Firstly, even traditional and introduced techniques of testing can be more or less altered and secondly, completely new and innovative testing formats may take a lot of effort when designing and a huge amount of time when going through the process of trial and error in practice. Therefore, Brown recommends to stay within the guidelines of accepted, introduced, traditional testing techniques (ibid.)

The first task when designing a reading test, according to Kalhous, is to ask ourselves what purpose the test is supposed to serve. Teachers may want to create summative tests at the end of term or year aiming to find out how their learners

understand the curriculum. They may also want to construct formative tests which are supposed to enhance learning and suggest ways to improvement (Kalhous 2002:220).

The next stage is, according to Brown determine appropriate objectives. He presents the situation on the example of learners who have been studying a unit with emphasis on the past. The objective of a reading test could therefore be focused on recognition of irregular past tense of selected verbs in a story describing an event in the past.

After stating objectives of the test, it is necessary to draw what specifications the test will have. The specifications later serve teachers as practical outlines of their tests, they include reading skills and sub skills to be tested and they state what topics you will cover, response formats for items, number of items, and length of the text. Length of texts may vary from very short poems to several pages long story. Longer texts, however, need to be divided into reasonably long sections. Exceeding reading time decreases concentration and attention and therefore influences the actual testing and grading respectively. The above mentioned limits may alter according to learners´ level of English, their age and other external influences (time of the day or day of the week).

Since chosen text has impact on the results, it is necessary to provide pupils with a sense of authenticity and interest and select such texts which will do so. However, according to Brown, when testing reading, authenticity may be endangered since texts may be edited to suit less advanced learners. We may then talk about simplified texts in terms of vocabulary and grammar. Nevertheless, he warns teachers that sometimes their effort to make texts as simple as possible may actually lead to the fact that these become a lot more difficult than the original text (Brown 1994:299).

Learners´ general knowledge goes hand in hand with text authenticity. ELT methodologists share the opinion that texts involving too much test-takers´ background knowledge should not be used very frequently (Hughes 1989:128). Kitao adds that the selected reading text must be the one which meets the requirement that the test-takers will have fairly similar levels of background knowledge on. In case there are pupils with deeper background knowledge, they will undoubtedly have a great advantage over those pupils who do not have it, regardless of their reading proficiency. At the same time, Kitao realises that it is rather difficult to construct such tests since it is quite impossible to predict how much learners already know about certain topic (Kitao&Kitao).

Next stage when creating a full scope reading test is, according to Brown, devising tasks for the test. These tasks may be written into a draft which serves as a review. When revising the draft, it is necessary to ask whether all the directions are absolutely clear, whether there is an example item for each task, whether the difficulty of each item is appropriate for the learners, whether the sum of the items and the reading test as a whole reflect the learning objectives, and finally whether each item is stated in clear and simple language. When constructing a reading test with multiple choice items, it is crucial to make sure that these items have appropriate distractors[7].

Once test-makers decide about test specifications, find a suitable text, and develop tests items, ELT experts recommend consulting the test with other professionals in the field who check its validity.

After experts have commented on the test's format, language, and items, it should be tried out on pupils who actually do not participate in taking the test so that the test maker may make any adjustments if they are needed (here, teachers may decide to involve a parallel class). However, as Brown suggests, the tryout phase is almost impossible in daily classroom teaching (Brown 2004:55).

Acts connected with preparation of the test, printing it, distributing it and so on are, according to Weir, parts of operation stage (Weir 1990:40). Moreover, teachers should make sure that their tests are neat and legible "so that all the effort put into construction of the test does not come to nought" (Brown 2004:55).

Another crucial parts of test construction procedure are decisions about scoring[8] and grading, and providing test-takers with feedback.

According to Brown, assigning grades to learner performance in a test is a very "thorny" issue and is influenced by many factors (country culture, institutional expectations, definitions of grades set in advance, or relationship with a particular class). At the same time, there are other factors which may influence teacher's grading, however, according to Gronlund, when grading a test it is necessary to:

> Base grades on student achievement and achievement only. Grades should represent the extent to which the intended learning outcomes were achieved. They should not be contaminated by student effort, tardiness, misbehaviour,

---

[7] See chapter 6 – Reading test techniques – Multiple-choice tests
[8] See chapter 6 – Reading test techniques

and other extraneous factors.... If they are permitted to become part of the test grade, the meaning of the grade as an indicator of achievement is lost.

(Gronlund in Brown 2004:284)

Feedback to students is, as Brown mentions, a very crucial part of the whole constructing tests process. Feedback that teachers offer therefore becomes beneficial washback. However, not every form of feedback is destined to be of positive character. Brown guarantees achieving positive washback if "students can, through the testing experience, identify their areas of success and challenge" (Brown 2004:63). Today, many teachers consider providing a number grade or a number of score when returning tests as one way of enhancing washback. In order to achieve positive washback, it is necessary to return the tests with narrative evaluations which give detailed descriptions of students' performance (praise of strengths or constructive criticism of weaknesses) and foster co-operative learning (Brown 1994:385-386).

Cyril Weir sees monitoring as the last stage of test construction process. He points out that in order to make tests as valid, reliable and efficient as possible it is necessary to monitor and store the results of each test. Then, these tests may serve as a mean of improving any new test in future (Weir 1990:40).

6. READING TEST TECHNIQUES

It is necessary to make clear distinction among the following three terms which may be easily confused – approach, method and technique.

When speaking about approach, ELT experts speak about a set of correlative assumptions dealing with the nature of language teaching and learning. Regarding methods[9], we speak about overall plan for the orderly presentation of language material, all of which is based on the selected approach. Thirdly, technique is a particular trick, strategy or contrivance used to accomplish an immediate objective. Furthermore, technique is always consistent with method and in harmony with approach.

Currently most widespread reading test techniques[10] will be discussed in this chapter with respect to their advantages, disadvantages and possible ways of scoring.

Weir claims: "It is possible to talk of a good or a bad test, valid or invalid test, but this is obviously not possible for test methods" (Weir 1990:42). The same opinion is shared by Alderson who declares that "there is no one best method for testing reading" (Alderson in Nuttall 2005:222).

Weir also suggests that each text is predestined for use of different test methods and Alderson is of the same opinion when he states:

> No single method can possibly be satisfactory for all the different purposes for which you might test, nor for all the skills and texts that you want to assess.
>
> (Alderson in Nuttall 2005:222)

Other ELT experts (Murphy, Porter, Boniakowska, and Alderson) share the opinion that test formats may actually affect performances of pupils. Similarly, Weir suggests that employing a range of valid, practical and reliable test formats is the only practical approach to preventing possible format effects (Weir 1990:42).

According to Madsen, reading tests may vary in terms of forms and they may evaluate a wide range of reading activities ranging from pre-reading activities, to reading speed, reading comprehension, scanning or skimming techniques, translation, and reading aloud, reading literature and so forth. However, it is **reading**

---

[9] Sometimes, ELT experts speak of ´methods´, (Alderson), sometimes methods are referred to as ´procedures´ (Nuttall)
[10] The term test ´technique´ may be used in the same meaning as test ´format´

comprehension that remains the most often evaluated reading activity in schools nowadays (Madsen 1983:76). According to Royer, reading comprehension is: "the process of understanding and constructing meaning from a piece of text. Phrases, sentences, paragraphs and so on are parts of connected text that can be read with comprehension" (Royer).

Walter claims that testing of reading is actually evaluating into what extent "the process of re-creating the text in the reader's mind" was successful. That is to say, testing of reading functions as feedback, since it involves responding to a certain text (Walter 1992:1).

ELT methodologists recognise two types of testing reading according to its production. First type is oral production and thus oral testing which concerns expressing pupils´ opinions, attitudes, ideas and feelings about the read material. The most common type of oral testing is retelling content of a text. Nevertheless, Heaton warns that in order to provide authenticity, teachers should assign tasks which pupils might encounter in real life and not to focus only on pure talking about a particular text (Heaton 1990:92).

Secondly, it is written production and thus written testing which will be discussed in this chapter.

When testing upper primary school grade pupils (eighth and ninth grade), it is possible to use many test techniques with respect to teacher's goals. According to Duffy and Roehler, "the upper primary school grades represent the bridging years between learning the fundamentals or reading and applying these fundamentals in a variety of specialized content areas" (Duffy and Roehler 2004:404). In other words, the goal shifts from pure print to reading comprehension and teachers guide reading using techniques that are more complex than the standard directed reading lessons. Nevertheless, we must bear in mind, that not every text is suitable for use of any test technique. Tests ideal for one purpose may be completely inadequate for other purposes and test format ideal in one situation may be of no use in other situations. Choice of the format depends on goals, purpose and importance of the test.

Speaking of goals emphasised with upper grade pupils, here are the most important ones, as Duffy and Roehler see them:

- Building vocabulary through direct study of words associated with content being studied

- Emphasis shifts from concrete words to multiple meaning words, homonyms, synonyms, antonyms, and other special categories of words
- Recognise fluently and instantly a wide variety of words
- Active prior knowledge of content using predicting
- Monitoring for unknown words in meaning and unrecognised words in print
- Recognise author's meaning
- Judge content of message by author's use of denotative and connotative word
- Use dictionary to find word meanings
- Develop a slow pace for careful reading and a fast pace for skimming
- Use summarising to remember content

(Duffy and Roehler 2004:405-407).

Reading tests may sometimes integrate testing language sub skills, in particular vocabulary and grammar. Heaton claims that "tests of vocabulary often provide a good guide to reading ability" since it is important to show not only knowledge of the meaning but also knowledge of collocations (Heaton 1990:79). When constructing a reading test that integrates vocabulary, it is necessary to decide which words students are expected to use (active or productive vocabulary) and which words they are expected to recognise (passive or receptive vocabulary) (ibid.).

The most widely spread formats of reading comprehension tests will be discussed on the following pages. However, it is necessary to mention that a test format considered used widely by one ELT expert often stays more or less unnoticed by other experts in literature.

6.1 MULTIPLE-CHOICE TESTS

According to Brown, the most crucial role of MCQs (multiple-choice questions) is testing reading comprehension. However, multiple-choice technique is widely used when testing reading knowledge of grammar and vocabulary as well. It is up to every teacher to decide into what extent context should be implied. They may test vocabulary or grammar tasks with a very little context provided in the text or they may test reading comprehension on more contextualised text (Brown 2004:194).

The following are only some examples why Weir believes that teachers are fond of MCQs. Firstly; tests of such type are completely reliably marked. That is to say that

there is no subjective evaluation since there are no alternatives to the correct answer. Furthermore, marking is simple and very effective in terms of time.

Secondly, MCQs are not confusing for test-takers because they know exactly what they are required to do in contrast to open-ended format.

Thirdly, multiple-choice tests do not require deploying of the writing skill unlike some other test formats (short answer questions for example) and therefore do not influence the skill being assessed (Weir 1990:43).

However, we may find many problems in connection with this technique. In the first place, there is much doubt about validity as a measure of language ability of such tests, since answering multiple-choice items is an unreal task, as in real life it is not probable that learners will be presented with four (three, five) alternatives of which only one is always correct.

For the second, devising satisfactory distractors may be time-consuming and very demanding (Weir 1990:43-44). Therefore, Heaton suggests that rather than creating a great number of absurd distractors (which every learner can see are wrong), it is preferable to select fewer options. However, these distractors must appear correct to anyone who is not sure of the correct answer (Heaton 1990:79-80).

Thirdly, Alderson sees a disadvantage in the fact that the distractors actually present learners with possibilities they may not otherwise have thought of (Alderson in Nuttall 2005:223).

The issue of MCQs tests scoring apart from being objective (thus reliable) may be rather questionable because there is always twenty-five percent (provided there are four options) probability of guessing the correct answer. Multiple-choice questions will be discussed in the practical part of the paper with respect to their scoring, effectiveness and formative role in teaching/learning processes.

6.2 TRUE/FALSE TESTS

True/false tests belong, according to Kitao, among the most common type of reading comprehension tasks.

She sees many advantages in employing them. To mention only some of them, they are quickly and easily constructed, they may be used even for very elementary learners and the true/false items can be written on a fairly short reading text which is

suitable even for less advanced readers (Kitao&Kitao). Another pro for employing this test format is, according to experts from UW Teaching Academy, the fact that the completed tests may be scored electronically (UW Teaching Academy).

On the other hand, many experts and teachers draw attention to the fact that this test technique encourages students to guess and thus gain score as high as 75 percent even though they only knew correct answers to fifty percent of the items. Methodologists suggest at least two ways of dealing with problems of guessing.

Madsen recommends introducing a penalty for guessing in form of awarding two points for each correct and one point deducted for each incorrect answer (Madsen 1993:88).

On the other hand, Kitao suggests providing a third alternative – ´not stated´ – and thus reduce the probability of guessing the correct answer. However, she is aware of the fact that sometimes it may be difficult to include a statement which seems close to the content of the reading passage and yet is not correct (Kitao&Kitao).

As stated above, many teachers and experts call attention to low validity of such test technique, however, there has so far been no evidence that this is true. Heaton is of different opinion. He claims that test-takers guess only if the test is too difficult for them and they are not sure of the correct answer. In order to prevent candidates from guessing, teachers must make the statements as understandable, clear, and concise as possible (Heaton 1988:84). True/false test format will be discussed in the practical part of the paper with respect to its scoring, effectiveness and formative role in teaching/learning processes.

6.3 MATCHING TESTS

Learners´ task is to match together appropriate items (synonyms, antonyms, interpretations of certain signs or labels) which are selected from a number of other options. This approach is, according to Alderson, useful for testing beginning learners (Alderson in Brown 2004:197).

As any other test technique, even matching has its pros and cons. Brown sees the biggest advantage in the fact that matching tasks offer an alternative to traditional multiple-choice format and are relatively easy to construct, as long as the matches have been selected carefully.

On the other hand, he points out the similarly to multiple-choice tests. Learners may never encounter matching tasks in real life situations. Brown´s advice is to provide more items so that the last pair is not so easy recognisable. On the other hand, he declares that sometimes this framework can "become more of a puzzle-solving process than a genuine test of comprehension as test-takers struggle with the search for a match, possibly among ten or twenty different items"(Brown 2004:198). The technique of matching will be further discussed in the practical part of the paper with respect to its scoring, effectiveness and formative role in teaching/learning processes.

6.4 PICTURE-CUED TESTS

According to Brown, there are two variations on this task. Firstly, learners are presented with a sentence or a short passage and choose one of a certain number of pictures being described (Brown 2004:199). Kitao adds that this type of task may be altered in such way that the candidates look at one picture, certain number of pictures and must decide which of the sentences accurately describes the picture (Kitao&Kitao).

Secondly, as Brown states, it is also possible to present test takers with a set of definitions or sentences. These describe a labelled part of a picture and the learners´ task is to identify each labelled item (Brown 2004:199). Picture-cued tests will be presented in the practical part of the paper with respect to their scoring, effectiveness and formative role in teaching/learning processes

6.5 ORDERING TESTS

Acquiring the skill of ordering and sequencing is necessary for remembering the order in which the author has arranged details of the text. Assembling little strips of paper in a logical story is a testing reading technique highly popular among both teachers and learners. Alderson, however, sees a danger in evaluating this test format. Often, there is not only one logical order of instructions, events, or ideas (Alderson in Brown 2004:209). An ordering test will be presented in the practical part of the paper with respect to its scoring, effectiveness and formative role in teaching/learning processes.

6.6 INFORMATION TRANSFER TESTS

According to Brown, information transfer test format is highly authentic and is essential for every educated person in order to be able to comprehend maps, charts, graphs, calendars, menus or class schedules. However, the sole comprehension of the above mentioned is not sufficient and must be accompanied by oral or written interpretations, comments, or questions. In other words, learners must imply a process of information transfer from one skill to another skill (from reading graphic or written information to speaking, for example) (Brown 2001:210).

6.7 SHORT-ANSWER TESTS

According to Brown, short-answer tests (sometimes referred to as open-ended questions) require the test-takers to produce answers based on information from the text in spaces provided on a partially completed chart. The candidates are presented with a set of questions which must be answered in one or two sentences (Brown 2004:207).

Tests of such format are most appropriate for questions that require student recall over recognition. In other words, the technique is employed in cases when it is desirable to ensure that the learners have committed the information to their memory (UW Teaching Academy).

The following are some advantages and disadvantages that experts from UW Teaching Academy, Weir, Brown, Heaton and Alderson see in use of this technique. The justification for this technique is that students are never provided with the right answer (as opposed to multiple-choice test) and therefore guessing is very much eliminated (Weir 1990:45). Brown mentions the fact that open-ended tests offer learners a chance to create their own answers which teaches them how to express themselves. Furthermore, there may be a possibility of potential follow-up discussion and therefore achieving beneficial washback effect (Brown 2004:207).

However, as any other test technique, even short-answer tests have their flaws. Firstly, the main disadvantage lies in the fact that this technique involves the candidate in writing and there is a possible danger of the teacher evaluating writing rather than reading skill. On the other hand, as Heaton suggests, accuracy of both written expression and reading comprehension need to be taken equally in account (Heaton 1993:87). Alderson also mentions that no short-answer test is easy to construct.

Furthermore, this test format is suitable only for questions that can be answered with short responses. According to him, teachers "must ask such questions that all possible answers are foreseeable otherwise they may be left with a bewildering range of responses" (Alderson in Nuttall 2005:223).

Scoring of such tests is very subjective, teachers are presented with variety of correct, partially correct, and incorrect answers and they must decide how to deal with them (Kitao&Kitao).

6.8 CLOZE TESTS

Nowadays, cloze tests are, without doubt, among one of the most popular types of reading assessment tests. The word was coined to capture "concept of ´closure´ that is, the ability to fill in gaps in an incomplete image (visual, auditory, or cognitive) and supply (from background schemata) omitted details (Brown 2004:201).

When constructing a cloze test, test-makers must decide if they want to delete every eighth word (experts recommend deleting of every fifth to twelfth words) regardless of whether this word is unfamiliar of whether it is a person's name. Methodologists then talk about ´fixed-ration deletion´.

On the other hand, if test designers use ´rational deletion´ (sometimes referred to as ´gap-filling procedure´) they omit words according to their discourse of grammatical functions. The only exception to this rule are the first two sentences which should stay without any blanks so that the test-takers can get used to the style and topic of the passage (Nuttal 2005: 222).

Scoring of cloze tests is, according to Brown, simple once the exact word scoring procedure is adopted. In other words, when teachers mark answers to this test, they can only give credit to test-takers only if they fill in the exact word which was deleted. However, there is another method of scoring, known as appropriate word scoring which can cause problems regarding face validity. That means that teachers can accept any word which is grammatically correct as long as contexts maintain sense (synonyms) (Brown 2004:201).

6.8.1 C-TESTS

As mentioned above, C-tests type of test is derivative of cloze tests. In other words, C-tests work on the same basis as standard cloze tests. Both methods work with deleted words. However, when creating a C-test, test-designers partially delete every second word in a text. To enable candidates to find solutions, they are presented with the first half of the deleted word (Nuttal 2005:224).

6.8.2 CLOZE ELIDE TESTS[11]

This technique is based on the fact that test-makers purposely insert grammatical or rhetorical errors in a text and the candidates must detect where these insertions have been made (Weir 1990:50).

According to Madsen, intrusive words must never be put in at fixed intervals. Words are added at random, however they must never come in succession (Madsen 1993:96). Even though cloze-elide technique is popular with many methodologists; Weir and Madsen mention a few disadvantages. Firstly, Weir points out that scoring is problematic to a certain degree since candidates may delete words which are redundant, yet correct (Weir 1990:50). Furthermore, experts claim that editing tests encourage word-by-word reading and thus essentially do not lead learners to the objectives of reading comprehension. What is more, use of this technique decreases reading speed (Madsen 1993:97).

6.9 SUMMARISING AND RESPONDING TESTS

Directions for writing a summary are quite simple. They usually include length and drawing attention to necessity to describe the main and supporting ideas of the text. Since evaluating summaries may sometimes be difficult for teachers, the following criteria serve as suggestions when assessing them. The ability to express accurately the main and supporting ideas is, according to Kitao, the most crucial one since it implies assessment of reading, and reading only (Kitao&Kitao). The other three factors (writing in own words, logical organisation of the summary, and clear expression of ideas)

---

[11] It is possible to encounter a term "editing test" in some methodology books (Madsen) or "intrusive word technique" (Nuttall).

pertain more to writing performance and it is up to the teacher's subjective decision how he or she evaluates the summary (Brown 2004:214).

According to Brown, responding to the text should not be confused with summarising as it often happens. The learner's task when writing a summary of a certain text requires expressing his on his or her opinion on the text. It is usual for the candidates to write into what extent they agree or disagree with the author and support their opinions with information from the text or from their own experience.

Similarly to summarising, scoring or responding tests may be rather subjective, depending on the accuracy of reflection of the text itself (ibid).

6.10 MIXED FORMAT TESTS

It is often possible to encounter test formats which are mixtures of some of the above mentioned techniques. For instance, learners may be presented with multiple-choice cloze, multiple-choice picture cued, or picture-cued true/false tests. Mixing techniques is, according to Brown, mostly used in nowadays testing (Brown 2004:192-193). Mixed format technique (ordering/picture-cued test) will be presented in the practical part of the paper with respect to its scoring, effectiveness and formative role in teaching/learning processes.

CONCLUSION OF THE THEORETICAL PART

To sum up the theoretical part, I strove to present a sufficient amount of well-founded and valid arguments on the benefits of integrating the skill of reading into lessons when acquiring a foreign language. It should be stated that reading is a necessary part of people's lives since a lot of real-life situations require people to employ their reading abilities in order to be successful in communication. Experts support the opinion that foreign language should be used the same way as a mother tongue, thus for communication. And communication indisputably involves not only the ability of saying or writing something, understanding other people's speech but also the ability to understand written texts. Even though it is not difficult to read in one's native language, reading in a foreign language may seem rather difficult, complex task requiring a great deal of practice and patience. Learners must be prepared for reading in real-life situations in which they will face authentic reading material. The practice in English lessons should provide sufficient amount of reading opportunities which will then help foreign language learners to become successful readers in their future lives.

Presenting learners with various reading activities that develop reading ability are likely to be of some interest and use is a demanding task for every teacher. Reading activities should be success-oriented so that they will not only increase motivation or ensure effectiveness of reading but will also build learners´ confidence in reading in foreign language.

Testing is an important part of everyday teaching and should not be passed by or even forgotten under any circumstances. Even though some teachers mistrust testing, well-made tests can actually help learners in their learning by establishing a positive attitude towards the studied language.

My thesis has suggested a framework for considering aspects which to focus on when designing tests with respect to reading in particular. It introduces important factors such as establishing purpose of a test, exploring nature of feedback that can be achieved, stating test specifications, selection of various reading genres, selection of test techniques, and possible ways of scoring reading tests.

What is taught is usually tested and on the contrary, learners usually study for what they know or at least sense is going to be tested. Therefore, great caution should

be taken when creating tests because only then can teaching become utterly effective. Tests which are labelled as effective must be valid, reliable, practical and they must provide positive washback.

I also strove to provide readers with enough background information so as they would be able to evaluate already existing reading tests with regards to their validity, reliability and practicality.

INTRODUCTION OF THE PRACTICAL PART

The objective of my theses is research into testing reading in EFL classes with focus on learners in upper primary school grades. I concluded that although reading and testing it can be employed even with beginner learners, teachers would more likely teach and subsequently test learners with higher knowledge of English. After considering the standpoint of level, I decided to choose nine grade pupils for my research.

The practical part deals with analysis of a set of reading tests of my own construction. The investigation is aimed at the aspects of the tests construction, with respect to the most widely used testing reading techniques, their effectiveness, scoring, and formative role in teaching/learning processes.

The analysed tests were based on six texts accompanying English language textbook English Project 2. Each testing included two tests based on a text. The testing was carried out during six months, considering the pace of one unit per month. Regarding the test formats, I strove to expose the pupils to the four most commonly used techniques of testing reading which have been suggested in the literature concerned.

As far as the content of the practical part is concerned, first of all, background of the research is described. After leading into my research, the particular tests are analysed individually. As I already mentioned, for the purpose of my research, I used a set of six reading tests, which were also examined in practice. In order to provide clear organization of the practical part and to make it explicit, I also submit the constructed tests after each analysis. However, samples of completed, scored and graded tests are placed in the appendix, alongside charts with scores of individual tests and the grading system. Furthermore, results of the presented test techniques are measured in terms of index of difficulty to see what the effectiveness of individual test technique is like. Data concerning difficulty indexes with respect to the test techniques are submitted in the appendix.

BACKGROUND OF THE RESEARCH

Firstly, when designing the tests, my first task was, according to literature studied; think of reasons for giving the tests. Since all the tests were written at the end of a unit, my reason for placing the tests was to assess pupils´ achievement over the period of approximately one month. However, I feel it is needless to say that my tests were constructed in order to improve learners´ reading ability rather than to test vocabulary or grammar.

The next step after having stated reasons for testing was to define the skills that I wished to test. All the tests were aimed at reading comprehension; therefore I wanted to employ learners´ skills which essentially led to the correct answers to comprehension questions or tasks. The skills and sub skills are all described in the test analyses.

The main idea of achievement reading tests is not to see how well learners can remember a known text, but how well they can read texts similar to those read in the course. The fact that texts had to be unknown for the pupils was assured by the fact that during our course, the Project English 1 texts were replaced by texts similar and later the original texts were introduced to the learners for the purpose of our testing. The texts used in the lessons during our course were ´borrowed´ from Project English 2, the textbook of English of later edition. The textbook was not used by pupils at the school; however, its content was very much similar to Project English 1 in terms of text genres, topics, vocabulary, and grammar. Therefore, when taking the tests, the learners were sure to be familiar with the nature of the test texts as far as their difficulty, length and other text specifications were concerned.

Having selected the texts and decided what skills I want to test, I needed to determine what method to use for testing comprehension of the texts. However, as already mentioned in the paper, according to ELT experts, "there is no best test method or technique since no method or technique can be suitable for all the different purposes for which we want to test, nor for all the skills and texts we want to assess" (Nuttall 2005:222).

Other decisions to be made were those concerned with response item formats and number of items. My intention was to use four most commonly reading test formats which would be suitable for the texts and for the skills I wanted to test. Regarding

number of items, I observed experts´ instructions that a test should number at least ten items in order to be considered sufficiently reliable.

Regarding time limits, after consulting literature concerned, I concluded that a twenty minute limit should be sufficient enough even for learners who have lower standard of English.

I also underwent the stage of discussing the tests with other teachers in order to eliminate any ambiguities or absurdities of the tests.

Another step was to make decisions about scoring. All the tests were scored objectively which is typical for testing reading. In order to eliminate guessing, weighing technique is introduced when scoring multiple-choice and true/false tests.

Many ambiguities still exist among test-makers as regards grading and there are several approaches to conversion of scores into grades. For the purpose of my research, I decided to choose the intuitive approach. In other words, my grading is fully subjective and I adjusted the criteria levels of performance according to my and my colleagues' judgements[12].

In order to ensure that my learners will be successful in passing the tests, I provided enough space for practising any issues that might be causing problems when writing the tests. Furthermore, the tests were always announced at least a weak in advance, so that the pupils had enough time to revise and eliminate any doubts concerning the subject matter.

I distributed the tests and always went through the instructions with the learners, so that the learners knew what was required. During the testing itself, the learners were allowed to ask questions in case there were any ambiguities as for the instructions.

As I already stated above, the time limit was twenty minutes on each test, nevertheless an extra minute was given to pupils with any problems.

Eventually, the tests were marked and returned as quickly as possible in order to provide a prompt feedback. Furthermore, it was possible to discuss specific points while the test was still in pupils´ minds. In conclusion, pupils were motivated to keep studying on regular basis.

---

[12] For grades see appendix

INDEX OF DIFFICULTY

As already mentioned above, one area of the test analyses concerns index of difficulty of the used test techniques. I would like to describe the method used for estimating index of difficulty and the steps, how the results were concluded. According to Kalhous, the extent to which a test is difficult is one of the main characteristics of a test. This value is measured according to the number of pupils who have answered a task correctly. For calculating index of difficulty of the test techniques, I used the instructions as Kalhous suggests them in literature (2002:228). First of all, all items of a certain technique in each test were added. Then the number of learners who participated in the test was located from the tests. Furthermore, it was necessary to find out what the number of pupils who have answered correctly is. Next step was to measure the index of difficulty according to this formula:

$$P = 100 \text{ x } (ns/n)$$

According to Kalhous, an ideal index of difficulty varies from $P = 20$ to 50. If the index falls below 20, he suggests that it is better to exclude the item from the test. Nevertheless, there is one option when we may leave such item in a test and that is if this item is the opening item of the whole text. He explains that the fact that the item is so easy to answer may lead to pupils´ relaxation (Kalhous 2002:229).

TEST ANALYSES

Reading test 1 – Unit 2

The test can be defined as achievement since it intends to measure the extent to which the students have mastered the material taught in the classroom. Furthermore, the test gives feedback on recent subject matter and encourages the learners in further studies. The skills that I wish to test are the ability to detect correct chronological sequence in the text and the ability to comprehend the text as a whole. The text and the tests are based on vocabulary and grammar contained in Unit 2 of English Project 1.

From the point of view of its construction, defined by Brown previously, it is a discrete test, since it isolates reading comprehension as the only aspect of language. According to Hughes, the test is indirect, because it measures the abilities which only underline the skills that we are interested in.

With reference to format, the test is divided in two parts, both tests applying a different testing technique.

The first exercise is focused on understanding given information and relating it to accompanying pictures by reordering accompanying pictures. Therefore, the mixture of ordering and picture-cued technique is used. Regarding scoring, it is objective, each item is awarded one point and there are no alternatives as to ordering the answers.

The second exercise focuses on learners´ ability to retrieve appropriate information from the text and true/false technique is used here. As far as scoring of this part of the test is concerned, the test is scored objectively since there are no alternatives to the correct answers. Moreover, weighting is applied in the latter mentioned exercise in order to discourage guessing. Weighting is done by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

Speaking about the instructions, they are written in English. Examples at the beginning of both tasks are not included because the instructions are clear and concise. Nevertheless, after distributing the copies, the learners were also given explanation of the instructions in Czech.

From the point of view of the scorer, I may say that marking was not time-consuming. For the purpose of practicality, each task contains the number of points achieved compared to the number of possible points.

Reading test 1 – Unit 2

Exercise 1 - Ordering and picture-cued test

Instructions: Read the original text and put the pictures in the correct order

The correct order of the pictures is:_____

/8

Exercise 2 - True/false test

Instructions: Read the original text and write down if the following sentences are T (true), F (false) or the text DS (doesn't say anything).

1/ You can make more than 12 cakes with these ingredients _____

2/ You bake the cakes for longer than one minute _____

3/ You put grated cheese under the potato cakes _____

4/ If you boil the potatoes for too long, they will not be nice _____

5/ Before you put the cakes on a tray, you must take put them in the oven _____

6/ Butter and margarine go in one bowl together with flour and the mashed potatoes _____

7/ It is fine if the potato cakes are black on both sides _____

8/ You need eight ingredients for baking the potato cakes _____

9/ You must slice the potatoes with a sharp knife _____

10/ You need less than half a kilo of potatoes for this recipe _____

/10

Total score   /18

Reading test 2 – Unit 3

The test can be defined as achievement since it intends to measure the extent to which the students have mastered the material taught in the classroom. Furthermore, the test gives feedback on recent subject matter and encourages the learners in further studies. The skills that I am interested in testing are the ability to find relevant information in the text and the ability to find information which is correct in the text. The text and the tests are based on vocabulary and grammar contained in Unit 3 of English Project 1.

From the point of view of its construction, defined by Brown previously, it is a discrete test, since it isolates reading comprehension as the only aspect of language. As far as Hughes's opinion is concerned, the test is indirect, because it measures the abilities which only underline the skills that we are interested in.

With reference to format, the test is divided in two parts both applying a different testing technique.

The first part aims to find out about the ability to relate information from the text in order to be able to answer questions; therefore multiple-choice technique is used here. Regarding scoring, it is objective, each item is awarded one point and there are no alternatives as to the correct answers. To discourage guessing, correction is introduced. It is made by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

The second part of the test aims to find out about the ability to relate information from one text to information from a different text and correct it. I opted for the use of matching technique here. As far as scoring of this part of the test is concerned, the test is scored objectively since there are no alternatives to the correct answers.

Speaking about the instructions, they are written in English. Examples at the beginning of both tasks are not included because the instructions are clear and concise. Nevertheless, after distributing the copies, the learners were also given explanation of the rubrics in Czech.

From the point of view of the scorer, I may say that marking was not time-consuming. For the purpose of practicality, each task contains the number of points achieved compared to the number of possible points.

Reading test 2 – Unit 3

Exercise 3 - Multiple-choice test

Instructions: Read the original text and circle the correct answer.

1/ The most important meal in Britain is

A/ breakfast   B/ lunch   C/ dinner

2/ What meal do people eat after noon?

A/ breakfast   B/ lunch   C/ dinner

3/ What meal was quite big in the past but is not very today?

A/ breakfast   B/ lunch   C/ dinner

4/ Which meal can people get in canteens?

A/ breakfast   B/ lunch   C/ dinner

5/ Which meal do people also call supper?

A/ breakfast   B/ lunch   C/ dinner

6/ Which meal do people prepare food from other countries for?

A/ breakfast   B/ lunch   C/ dinner

7/ Which meal is quite big at weekends?

A/ breakfast   B/ lunch   C/ dinner

8/ Which meal has ham, cheese or vegetable between two slices of bread?

A/ breakfast   B/ lunch   C/ dinner

/8

Exercise 4 - True/false test

Instructions: Read the original text and correct the false sentences. Cross out the wrong word and write the correct one. Some of the sentences are correct. If you find a correct sentence, write C (correct) next to it.

1/ Lunch is the most important meal of the day.   _____

2/ British people often eat sauces in lunchtime.   _____

3/ Many people do not eat meat.   _____

4/ British people usually prepare big breakfast only on Saturdays.   _____

5/ There are toasts in a typical British lunch.   _____

6/ Children can have their lunch at school or at home.   _____

7/ A lot of people like to cook food from other countries.   _____

8/ Grapes are peoples´ favourite fruit. _____

9/ It is quite usual to have something sweet after dinner. _____

10/ Food from China or India is very popular nowadays. _____

11/ In some parts of Britain, ´dinner´ means the same thing as ´breakfast´. _____

/11

Total score /19

Reading test 3 – Unit 4

The test can be defined as achievement since it intends to measure the extent to which the students have mastered the material taught in the classroom. Furthermore, the test gives feedback on recent subject matter and encourages the learners in further studies. The skills I want to test are the ability to find relevant information in the text and the ability to recognize and interpret cohesive devices within a text. The text and the tests are based on vocabulary and grammar contained in Unit 4 of English Project 1.

From the point of view of its construction, defined by Brown previously, it is a discrete test, since it isolates reading comprehension as the only aspect of language. As far as Hughes's opinion is concerned, the test is indirect, because it measures the abilities which only underline the skills that we are interested in.

With reference to format, the test is divided in two parts, both tests applying a different testing reading technique.

The first exercise focuses on retrieving appropriate information from the text and multiple-choice questions technique is used here. Regarding scoring, it is objective, each item is awarded one point and there are no alternatives as to the correct answers. To discourage guessing, correction is introduced. It is made by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

The sub skill tested in the second exercise is the ability to interpret pro-forms. The relationship is signaled in the text by anaphoric or cataphoric reference. In order to interpret pro-forms (meanings of words like "it, our, them, he, his"), I opted for the use of matching technique here. As far as scoring of this part of the test is concerned, the test is scored objectively since there are no alternatives to the correct answers.

From the point of view of the scorer, I may say that marking was not time-consuming. For the purpose of practicality, each task contains the number of points achieved compared to the number of possible points.

Reading test 3 – Unit 4
Exercise 5 - Multiple-choice test
Instructions: Read the original text and circle the correct answer
1/ Russell never sends Valentine cards

A/ because he is very lazy   B/ because he doesn't like any girl in his class

C/ because he is very popular with girls in his class

2/ Greg is in love with Nicola

A/ because she is very kind and pretty   B/ because she likes him too

C/ because she sits near him at class

3/ Greg bought the Valentine card

A/ because Peter liked it too   B/ because he thought that the words in

very nice   C/ because it was cheaper than other cards

4/ Greg wanted to take the card from Nicola´s desk

A/ because he didn't put down his name   B/ because Amy saw him putting it

there   C/ because he heard that Nicola has a card for Russell

5/ Greg, Peter, Nicola, Amy, Russell and other children went out of the classroom

A/ because the school finished that day   B/ because they had free time between

two lessons   C/ because they went to the school canteen

6/ Greg was very happy when he came back to the classroom

A/ because he really liked what Nicola did to Russell   B/ because he talked to

Nicola outside   C/ because Nicola likes him too

/6

Exercise 6 – Matching test

Instructions: Read the original text, then read the sentences and replace the underlined pronouns with the nouns from the list. You can use some of them more than one time.

Russell   Greg   Nicola   cards   the card   Amy   Peter

1/ They left the building after lesson – _____

2/ She really liked what Nicola did to Russell - _____

3/ It cost more money than the rest – _____

4/ They entered the classroom together – _____

5/ He reminded him of the St. Valentine's Day - _____

6/ Everyone made fun of him – _____

7/ He wasn't happy at all - _____

8/ I will receive a great number of them – _____

9/ He was very popular with girls – _____   /9

Total score   /15

Reading test 4 – Unit 5

The test can be defined as achievement since it intends to measure the extent to which the students have mastered the material taught in the classroom. Furthermore, the test gives feedback on recent subject matter and encourages the learners in further studies. The skills I wish to test are the ability to understand important information mentioned in the text. The text and the tests are based on vocabulary and grammar contained in Unit 5 of English Project 1.

From the point of view of its construction, defined by Brown previously, it is a discrete test, since it isolates reading comprehension as the only aspect of language.

According to Hughes, the test is indirect, because it measures the abilities which only underline the skills that we are interested in.

With reference to format, the test is divided in two parts, both tests applying a different testing reading technique.

The first exercise is focused on understanding relations mentioned in the text. Therefore, I chose to use matching technique. According literature, it is advisable to provide more items so that the last pair is not so easy recognisable. Regarding scoring, it is objective, each item is awarded one point and there are no alternatives to the correct answers.

The second exercise is focused on learners´ ability to find certain information in a map. Therefore, information transfer technique was used here. As far as scoring of this part of the test is concerned, the test is scored objectively since there are no alternatives to the correct answers.

Speaking about the instructions, they are written in English. There are no examples at the beginning of both tasks because the instructions are clear and concise. Nevertheless, after distributing the copies, the learners were also given explanations in Czech.

From the point of view of the scorer, I may say that marking was not time-consuming. For the purpose of practicality, each task contains the number of points achieved compared to the number of possible points.

Reading test 4 – Unit 5

Exercise 7 - Matching test

Instructions: Read the original text and match the sentences. Write the letters in the gaps provided. You do not need to use all the answers.

1/ It is located in California   ____

2/ It is the year when USA got its independence   ____

3/ It is the population of the country   ____

4/ It is a building where the US president lives   ____

5/ It is the third longest in the world   ____

6/ It is on the south boarder of the USA   ____

7/It is a little state in the east of USA   ____

8/ It is the biggest city in the country   ____

9/ It is the fourth largest in the world   ____

10/ It is on the northern boarder of the USA   ____

A/ the United States of America

B/ the Mississippi-Missouri River

C/ New York

D/ 8 million

E/ Canada

F/ Chicago

G/ Los Angeles

H/ 260 million

I/ Washington D.C.

J/ Mexico

K/ the White House

L/ California

M/ 1776

N/ Florida

O/ Delaware

/10

Exercise 8 - Picture-cued test

Instructions: Look at the map and match the following descriptions and the geographical names from the map. Write them in the gaps provided.

1/ The American state on the west coast of the USA - _____

2/ The country bordering with the USA in the north - _____

3/ One of the small American states above New York - _____

4/ The city on the coast of the Great Lakes - _____

5/ The most southern state of the USA - _____

6/ The country bordering with the USA in the south - _____

7/ The longest river in the USA - _____

8/ The state between two cities - _____

9/ The city under Delaware - _____

10/ The state on the south east coast of the USA - _____

12/ The biggest city in the west coast of the USA - _____

/12

Total score   /22

Reading test 5 – Unit 6

The test can be defined as achievement since it intends to measure the extent to which the students have mastered the material taught in the classroom. Furthermore, the test gives feedback on recent subject matter and encourages the learners in further studies. The skills that I am interested in testing are the ability to understand relations between parts of a text and the ability to retrieve specific information from the text and accompanying pictures. The text and tests are based on vocabulary and grammar explained in Unit 6 of English Project 1.

From the point of view of its construction, defined by Brown previously, it is a discrete test, since it isolates reading comprehension as the only aspect of language. According to Hughes, the test is indirect, because it measures the abilities which only underline the skills that we are interested in.

With reference to format, the test is divided in two parts, both tests applying a different testing reading technique.

The first exercise focuses on learners´ ability to recognize text organisation and insert parts of a text in correct places. Pupils´ task is to complete the correct key sentences in gaps provided, therefore matching technique is used here. As literature suggests, I provided more items so that the last pair was not so easy to recognise. Regarding scoring, it is objective, each item is awarded one point and there are no alternatives to the correct answers.

The second exercise is focused on learners´ ability to comprehend statements from the text and compare them with pictures therefore picture-cued technique is used. As far as scoring of this part of the test is concerned, the test is scored objectively since there are no alternatives to the correct answers. On basis of literature, I again opted for providing more items so that the last match is not so easy to recognise.

Speaking about the instructions, they are written in English. Examples at the beginning of both tasks are not included because the instructions are clear and concise. Nevertheless, after distributing the copies, the learners were also given explanation of the rubrics in Czech.

From the point of view of the scorer, I may say that marking was not time-consuming. For the purpose of practicality, each task contains the number of points achieved compared to the number of possible points.

Reading test 5 – Unit 6

Exercise 9 - Matching test

Instructions: Fill the sentences (A-G). Write the letter in the gaps provided. There are three extra sentences that you do not need.

Luke Ryan opened his eyes slowly. He was on a bed in a strange room. Kirsty was there, too. "Are you all right?" she asked quietly. "It doesn´t matter. You´re both going to die." It was Luke´s old enemy, Doctor Dread. He laughed horribly. "I´ve got the computer chip now. So goodbye, Mr Ryan." _____

"It looks like the end," said Kirsty sadly. "No, we must stop him," said Luke. "With that chip he can destroy all the world's satellites." "I know," said Kirsty. "That guard is very slow. I think we can escape easily." She pointed to the blanket on the bed and whispered to Luke. "That's a good idea," said Luke and he started to shout very loudly. "Oh, my stomach. Help! My stomach! Help!"

_____"What's all that noise?" he said. Kirsty threw the blanket over the guard´s head. Luke hit him hard in the stomach and quickly grabbed his gun and the keys. "You did that very well," said Luke, as they locked the guard in the room and ran out. "Oh, it was easy," said Kirsty.

_____Luke pushed the driver out and they drove as fast as possible to the airport. When they arrived at the airport, they saw a small plane. "That's Dread's plane," said Luke. They ran to the plane. It was empty, so they got in and hid behind the seats. Soon Doctor Dread got into the plane. _____

_____ "With that chip I'm going to be master of the world," he said proudly. "No, you aren't," shouted Luke. He jumped up from behind the seat and grabbed the bag. "Give that to me," said Doctor Dread. He quickly grabbed the bag and pulled hard. But he pulled too hard. The door behind him suddenly opened and he fell out. With the bag in his hand he fell down and down and down into the sea. _____

A/ There was a car in the street.

B/ Luke and Kirsty opened the door.

C/ The car was very small.

D/ Kirsty flew the plane back to the airport and they landed safely.

E/ He left the room and the guard locked the door.

F/ The guard rushed in.

G/ He carefully put a bag on the seat next to him, started the plane and took off.

H/ Ten minutes later, when they were over the sea, he looked at the bag.

/6

Exercise 10 - Picture-cued test

Instructions: Look at the pictures in the original text and match them with these sentences. There is one sentence that you do not need to use.

1/ The guard looked in the room through the little door window _____

2/ Luke grabbed the bag _____

3/ Luke drove the car to the airport _____

4/ Luke and Kirsty threw a blanket over the guard's head_____

5/ The door opened and Dr. Dread fell out of the plane _____

6/ Dr. Dread visited Luke and Kirsty in their room _____

7/ Kirsty had an idea and she told Luke about it _____

8/ Kirsty and Luke got in the plane _____

9/ Luke and Kirsty watched Dr. Dread as he fell out of the plane _____

10/ Kirsty drove the car to the airport _____

/9

Total score   /15

Reading test 6 – Unit 7

The test can be defined as achievement since it intends to measure the extent to which the students have mastered the material taught in the classroom. Furthermore, the test gives feedback on recent subject matter and encourages the learners in further studies. The skills that I am interested in testing are the ability to interpret lexical cohesion and the ability to correct wrong lexical cohesion in forms of using synonymy. The text and tests are based on vocabulary and grammar explained in Unit 7 of English Project 1.

From the point of view of its construction, defined by Brown previously, it is a discrete test, since it isolates reading comprehension as the only aspect of language. As far as Hughes´s opinion is concerned, the test is indirect, because it measures the abilities which only underline the skills that we are interested in.

With reference to format, the test is divided in two parts, both tests applying a different testing reading technique.

The first exercise is focused on learners´ ability to find the same lexical items to refer to one and the same thing and therefore, multiple-choice technique is used here. Scoring of this test is objective, each item is awarded one point and there are no alternatives as to ordering the answers. Moreover, weighting is applied in the latter mentioned exercise in order to discourage guessing. Weighting is done by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

The second exercise focuses on the ability to find the same lexical items to refer to one and the same thing, only this time; learners must find the correct synonym in the text and correct the wrong sentences. I opted for the use of true/false technique here. As far as scoring of this part of the test is concerned, the test is scored objectively since there are no alternatives to the correct answers.

Speaking about the instructions, they are written in English. Examples at the beginning of both tasks are not included because the instructions are clear and concise. Nevertheless, after distributing the copies, the learners were also given explanation of the rubrics in Czech.

From the point of view of the scorer, I may say that marking was not time-consuming. For the purpose of practicality, each task contains the number of points achieved compared to the number of possible points.

Reading test 6 – Unit 7

Exercise 11 - Multiple-choice test

Instructions: Read the original text and choose the correct answer

1/ The type of entertainment that people like the most in Britain is.....

A/ cinema   B/ going to the library   C/ TV

2/ Usual channels in Britain put on ....programmes

A/ various   B/ same   C/ similar

3/ The channels which are not free put on .... of programmes

A/ a similar type   B/ various types   C/ the same type

4/ Videos are favourite with people because..........

A/ are cheaper than cinema tickets   B/ people like sitting at their homes not in cinemas

C/ a video shop is in almost every town but cinemas only in large towns

5/ When you go to the theatre, you can see ....

A/ people whose job is acting   B/ people whose hobby is acting   C/ both types of people

6/ If you see a pantomime in the theatre, it will be................

A/ very sad   B/ very funny   C/dramatic

7/ If they lend you books in a public library, you ....

A/ should pay   B/ don't have to pay   C/ must pay

/7

Exercise 12 - True/false test

Instructions: Read the original text and correct the false sentences. Cross out the wrong word and write the correct one. Some of the sentences are correct. If you find a correct sentence, write C (correct) next to it.

1/ Television is the least favourite type of entertainment in the UK.   _____

2/ They have a theatre in every large town.   _____

3/ There is a private library in all towns and a lot of villages. _____

4/ Most theatres show a comedy in November and December. _____

5/ Digital and satellite channels are not free in the UK. _____

6/ Children can watch children shows from four o'clock to six thirty. _____

7/ People can buy videos in video shops. _____

8/ If you want to borrow something to read in a library you don't have to pay. _____

/8

Total score   /15

CONCLUSION OF THE RESEARCH

The results of the research show that testing is an inevitable part of teaching process. This is supported by washback effect. If teachers consider reading a full scope language skill, then it is worth including it in pupils´ assessment to build positive attitude towards reading. On the contrary, if reading is practised in lessons, but very rarely tested, learners may come to conclusion that reading does not really matter because it is not worth being tested.

Testing reading is a more complicated issue than may be at first assumed. A great number of important factors need to be taken in consideration in development of a well-constructed reading test.

The theoretical part of the paper provides readers with enough information in order to be able to evaluate already existing reading tests from the point of view of their validity, reliability and practicality. Nevertheless, the main objective of the thesis is to give a framework for constructing one's own classroom tests with respect to testing techniques, their effectiveness, scoring and their formative role.

The fact that the purpose of a test needs to be set before any choice of a particular test format can be made is probably the greatest lesson to be learnt from my research. What is more, even before any test construction it is necessary to explore the nature of feedback that can be achieved from the test. Subsequently, feedback on the effectiveness of pupil learning is usually of interest to both learners and teachers. In addition, this information is of use for the teacher if he or she wants to apply formative assessment.

When the test purpose has been stated, it was necessary to consider test specifications, skills and sub skills we want to test, a suitable text, test format, number of items and scoring. Allocating numbers of points directly in rubrics proved to be practical and after setting criteria for contingent grading would become fast and not time-consuming.

As far as the test techniques are concerned, reading offers a wide range of possibilities. Comparing results of the tests in terms of difficulty of the tests (test techniques respectively), it can be concluded that the least difficult and therefore most suitable technique for assessing reading comprehension is multiple-choice. Many things

speak in beneficial use of this technique, for example the fact that its clear instructions were never confusing for learners. Secondly, the format of answering MCQ tests was not demanding for the learners in terms of time and writing. However, after distribution of test introducing MCQ technique, I noticed that many children skipped reading the text altogether and went directly to answering the test items. They first read the item and then skimmed over the text until a response was located, repeating the process for all items. Their activity was more aimed at locating information rather than constructing meaning form the texts. Therefore, it can be concluded, that the tests may have been effective in producing high test scores, however they only reflected little engagement with the texts. To conclude, MCQ is a reliable test format relatively easily constructed and objectively and effectively scored, not confusing nor demanding for learners in terms of answering, however, validity of this test format (are the tests really measuring reading comprehension?) is rather disputable.

Another successful test technique according to the test results was matching technique. Matching is a similar technique to multiple-choice technique and therefore many of the same rules apply. Like MCQs, the test format is not realistic which may have decreased learners´ motivation to a certain degree. The suggestion to provide more detractors than needed proved to be very useful and eliminated arriving at a correct answer through a process of elimination, however, during writing the tests, I sometimes noticed that learners were sometimes puzzled by these detractors. When scoring the test, I could see that I did not provide the pupils with all necessary instructions because some of the test-takers have used some of the test items more than one.

Due to their relatively low index of difficulty, the true/false tests proved that the items must be constructed very carefully in order to be correctly answered. It was necessary to design such statements which were written in simple, direct language and did not confuse learners and did not promote guessing. Furthermore, it was crucial to provide statements which were paraphrased, and not directly quoted which may have caused some of the wrong answers because the learners did not understand the test items. When scoring the tests, I also noticed that there was some ambiguity among pupils concerning the fact, that some of the pupils made a mistake of considering parts of false sentences true, although there was a false statement. On the other hand, what surely increased validity of the test format was the fact that learners were not exposed

to any long statements or relative and qualifying words restricting or opening general statements (usually, sometimes and so on). In order to test reading comprehension in greater extent and to eliminate guessing, in some exercises, students were asked to correct false statements and a third option "doesn't say" was added which, I think, may have lead to relatively low number concerning the index of difficulty.

Picture-cued technique proved to be the most demanding technique for my pupils according to the results. There are many things favouring this test format (it is more realistic than other reading test formats, for example MCQ or matching and thus comes much closer to the real world which motivates learners positively. Furthermore, it spares learners from reading difficult and many times rather dull texts). For learners in upper primary school grades, however, use of this technique may be rather disputable since employing pictures which provide information requires only little or even no reading.

To conclude, I feel that it is necessary to say that there is no best method, technique or test format. A test technique which is ideal for one purpose may be of no use for another purpose; a test format which may work very well in one situation, may be inappropriate in another situation. The choice of test technique is dependant on the course objectives, the purpose and the importance of the test and available resources.

Samples of completed tests, charts with results and grades and data for measuring index of difficulty are included in appendix.

RESUMÉ

Ve své diplomové práci se zabývám tématikou testování čtení na základních školách. Můj výzkum je zaměřen na testovací techniky vhodné pro žáky devátého ročníku, kteří pracují s učebnicí English Project 1. Pro praktickou část své diplomové práce jsem vytvořila sadu šesti testů, které byly následně ověřeny i v praxi. Výsledky šetření, včetně hodnocení využitých testovacích technik, jsou uvedeny v závěru práce.

Teoretická část je rozdělena do šesti kapitol. Každá z těchto kapitol se zabývá různými aspekty týkající se problematiky testování. První tři kapitoly se věnují obecným pojmům z oblasti testování, zatímco zbylé tři kapitoly jsou zaměřeny na testování jazykové dovednosti čtení.

První kapitola upozorňuje čtenáře na velmi blízký vztah mezi testováním všeobecně a výukou, jelikož testování je považována za neodmyslitelnou součást výuky. Vedle toho též plní důležitou roli v procesu evaluace. Tři spolu úzce spjaté pojmy, tj. evaluace, hodnocení a testování, jsou objasněny v rámci první kapitoly. Evaluace je totéž, co výsledek hodnocení, tedy vyhodnocení. Naopak, pokud máme na mysli hodnocení ve smyslu "assessment", tento proces se zabývá především měřením výkonů žáka a získáváním hodnotících informací ve výuce. Motivační, poznávací a konativní jsou tři hlavní funkce hodnocení. Motivační funkce souvisí s emocionální, jinými slovy citovou stránkou hodnocení, zatímco poznávací funkce je zaměřena na intelektuální, tedy rozumovou stránku hodnocení. Konativní funkcí se rozumí, že hodnocení působí na skutečnost, jinými slovy souvisí s vůlí člověka k činu. Hodnocení se dále dělí na analytické a holistické. Analytické hodnocení, které je záměrné, je možné bez problémů analyzovat a formalizovat, což nelze u holistického hodnocení, které je bezděčné. Výsledkem takovéhoto formalizovaného hodnocení je známkování. Z jiného hlediska je hodnocení rozděleno na sumativní a formativní. Smyslem sumativního hodnocení je získat celkový přehled o dosažených výkonech, na formativní hodnocení je nahlíženo jako na korektivní či zpětnovazební. V neposlední řadě Slavík popisuje také normativní a kriteriální hodnocení.První kapitola se dále zabývá efektem, které na výuku testování má. Tento vliv, který může být jak pozitivní tak negativní byl do českého jazyka přijat pod pojmem "backwash" neboli "washback". V neposlední řadě první kapitola pojednává o mnoha funkcích, které testy mají. Jednou z těchto funkcí je

to, že testování slouží učiteli jako zpětná vazba pro případné zlepšení výuky a způsobu hodnocení. Testování zajišťuje zpětnou vazbu i pro žáky, například tím, že odhalí případné nedostatky a potíže s učivem. V neposlední řadě je testování prostředkem motivace žáků k učení a hodnotícím nástrojem výkonů, tedy zdroj informací pro evaluaci.

Klasifikaci různých typů testů a postupů přípravy testů je věnována druhá kapitola. Z hlediska účelu testu je možné rozlišit pět typů testů: testy výkonové (achievement), testy úrovně (proficiency), testy diagnostické (diagnostic), testy zařazovací (placement) a testy vhodnosti (aptitude). Výkonové testy jsou vázány na probíranou látku a jsou odlišené podle časového zařazení do výuky. Testy průběžné, vycházející z krátkodobých cílů (progress achievement) a testy výstupní (final achievement), které jsou zařazeny na konci pololetí, školního roku nebo na konci kurzu. Testy úrovně mají za cíl zjistit úroveň jazykových znalostí bez závislosti na předešlé výuce. Diagnostický test je používán k objevení znalostí a nedostatků, z kterých lze určit další postup při výuce. Test zařazovací slouží k přiřazení studentů do příslušné úrovně výuky. Posledním typem testů jsou typy vhodnosti, které odhadují do jaké míry je žák vybaven schopnostmi naučit se cizí jazyk. Z hlediska postupu přípravy testů je testování možné dělit na testování přímé (direct) a nepřímé (indirect). Přímé testování vyžaduje konkrétní dovednost (například čtení či psaní), zatímco nepřímé testování se zaměřuje na určitou strukturu. Dalším typem je testování jednotlivé (discrete point), které testuje jednu dovednost oddělenou od ostatních a integrující (integrative point), které tyto dovednosti naopak spojuje. Podle hlediska interpretace žákova výkonu se testy dělí na rozlišující a ověřující. U prvně zmíněného typu se výsledek testu posuzuje vzhledem ke všem testovaným. Na druhou stranu, u testů ověřujících se výsledek posuzuje vzhledem k dosažené úrovni jazyka. Z hlediska míry objektivnosti skórování rozlišujeme testy objektivně a testy subjektivně skórované.

K tomu, aby test byl dostatečným prostředkem měření výsledků výuky, je potřeba, aby měl jisté vlastnosti. Validita, reliabilita a praktičnost jsou základními vlastnostmi každého dobrého testu a tyto jsou popsány v druhé kapitole teoretické části. Test splňuje podmínky validity, pokud se jím zkouší to, co má opravdu být zkoušeno. Je možné rozlišit několik druhů validity, například obsahovou nebo konstruktivní. Reliabilita testu je zabezpečena tehdy, jsou-li výsledky minimálně dotčeny náhodnými

vlivy. Poskytuje-li test stejné výsledky za stejných podmínek, je možné hovořit o spolehlivém, tedy reliabilním testu. Přesnost testu je další podmínkou pro vysokou reliabilitu. Jinými slovy, jestliže při použití testu nedochází k velkým chybám měření, je test přesný. Reliabilitu testu je možno měřit různými metodami, např. metodou půlení, která je použita v praktické části práce. Podmínkou toho, aby byl test dostatečně validní je, že musí mít vysokou reliabilitu. Na druhou stranu, vysoká reliabilita testu nezaručuje, že test je opravdu validní. Test může spolehlivě a přesně měřit určité vědomosti i tehdy, jestliže měří něco zcela jiného, než měřit má. Praktičnost je další kritérium, na které je při sestavování testů vedle validity a reliability kladen důraz. Je-li použití testu snadné, stejně tak jako oprava jeho výsledků, lze říci, že test je praktický. Na závěr lze konstatovat, že reliabilní a validní test je v praxi těžko použitelný, pokud zároveň není praktický.

"Je všeobecně známo, že čtení v mateřském jazyce pomáhá rozšiřovat slovní zásobu a přispívá ke schopnosti vyjádřit sebe sama, protože mnohem více slov je užíváno v psaném nežli v mluveném projevu" (můj překlad, Brown: 1994, s.290). Toto Brownovo tvrzení zdůrazňuje význam čtení, které je tématem čtvrté kapitoly. Čtení je mnoha učiteli považováno za samozřejmé, protože většina lidí ve vyspělém světě se naučí číst v ranném věku. Nicméně je nezbytné uvědomit si, že čtení v cizím jazyce není tak přirozené jako čtení v mateřském jazyce a tudíž by mu žáci neměli být vystaveni aniž by byli na ovládnutí této jazykové dovednosti systematicky připravováni. Tématem této kapitoly jsou mimo jiné i různé žánry textů, které je možné použít pro testování. Pozornost je dále věnována i mikro a makro dovednostem, spojených s testováním čtení. Co se týče strategií uplatněných při této jazykové dovednosti, jsou podrobněji prozkoumány strategie "skimming" a "scanning". Tyto dva pojmy byly z anglického prostředí převzaty do češtiny a odpovídají rychlému přečtení textu za účelem zjištění tématu textu, jeho účelu a organizaci, respektive nalezení odpovídajících informací v textu, aniž by byl tento celý přečten. Z hlediska rozlišení typů čtení, můžeme hovořit o čtení nahlas a čtení tichém. Mezi odborníky je v současné době pro testovací účely upřednostňováno spíše tiché čtení a to z mnoha důvodů. Jedním z nich je například i fakt, že při čtení nahlas se testovaný soustředí na správnou výslovnost a to může mít za následek sníženou schopnost vnímat smysl textu.

Dalším hlediskem pro rozlišení typů čtení je intenzita, s jakou je daný text čten. Hovoříme-li o intenzivním čtení, máme na mysli hluboké a důkladné vnímání textu, zatímco extenzivní čtení slouží jako termín pro obecné porozumění a zahrnuje delší než jednostránkové texty.

Pátá kapitola se zabývá popisem jednotlivých fází přípravy testů čtení. Nejprve je nutné zamyslet se nad tím, jaký vlastně je účel testu a co se od něho chceme dozvědět. Poté je potřeba určit jaká specifika bude test mít. Tato specifika určují, na jaké mikro a makro dovednosti se chceme v testu zaměřit, určují téma textu, testovací techniku, počet odpovědí a také délku textu. Při výběru textu je dobré dbát na autenticitu a poutavost textu. Důležitým kritériem při výběru textu je i fakt, do jaké míry jsou žáci při testu ovlivněni základní znalostí. Další fází při přípravě testů je samotné sestavení otázek. Neméně důležitou fází je konzultace testu s kolegy za účelem odstranění všech mnohoznačností a nepřesností. Každý test by měl projít jakousi "zkouškou", například v paralelní třídě, aby mohly být učiněny úpravy, jsou-li nutné. Rozhodnutí ohledně skórování, známkování a poskytnutí zpětné vazby žákům patří také mezi fáze přípravy testů. Chceme-li, aby zpětná vazba byla pro žáky pozitivní, nestačí pouze napsat známku a počet dosažených bodů. Naopak, testy by měly být vraceny s podrobným popisem, jinými slovy pochvalou žákových silných stránek a konstruktivní kritikou jeho slabin. Známkování testů je ovlivněno mnoha faktory působící na učitele, jako například kultura země, výsledky předešlých testů nebo sympatie či antipatie k určitému žáku či třídě. Nicméně je důležité, aby známky odrážely pouze a jenom žákův výkon.

Výběru testovací techniky by měla být věnována zvláštní pozornost, jelikož může ovlivnit jak validitu testu, tak výkon žáka. Proto odborníci doporučují využití mnoha validních, reliabilních a praktických testových formátů. Lze testovat mnoho čtecích aktivit počínaje rychlostí čtení, přes překlad, hlasité čtení až k porozumění textu. Právě porozumění textu je dnes nejčastěji testovanou aktivitou a má dvě podoby. Jedná se o ústní testování, kdy žák vyjadřuje své názory či pocity o daném textu, nebo pouze převypráví obsah. Na druhou stranu existuje testování zahrnující písemné odpovědi, a právě touto formou testování se zabývá praktická část mé diplomové práce. Mezi techniky písemného testování čtení lze zařadit: testové úlohy s výběrem odpovědí (multiple-choice), přiřazující úlohy (matching), úlohy přiřazující s použitím obrázků

(picture-cued),dichotomické testové úlohy (true/false), úlohy s otevřenou odpovědí (short-answer), testové úlohy doplňovací (cloze, cloze elide, c-tests), úlohy s přenosem informací (information transfer), úlohy uspořádací (ordering), úlohy shrnující (summarising) a v neposlední řadě lze také kombinovat formáty určitých testů, například úlohy s výběrem odpovědí a úlohy doplňovací (multiple-choice cloze).

Praktická část mé práce je založena na výzkumu, který byl uskutečněn v devátém ročníku základní školy. Sada šesti testů různých testových technik byla vytvořena pro žáky ve věku čtrnácti či patnácti let. Uvedené testy byly založeny na základě učebnice Project English 1 používané v hodinách anglického jazyka. Velice důležitým faktorem bylo, že žáci nesměli být testováni na základě známého textu. Proto byly použity texty z učebnice novějšího vydání učebnice s názvem Project English 2. Texty byly textům z probírané učebnice podobné, co do podobnosti žánru, tématu, délky, obtížnosti a také slovní zásoby a gramatických jevů. Testy na základě šesti textů byly mnou vytvořeny a při respektování testovacích podmínek použity v hodinách anglického jazyka a následně vyhodnoceny. Ve výzkumné části práce je analyzováno celkem šest testů z hlediska jejich přípravy, respektive použitých testovacích technik. Použité testy jsou pro přehlednost zařazeny k jednotlivým analýzám, vyplněné testy a údaje o jejich hodnocení jsou uvedeny v příloze. Vlastní testování se uskutečnilo šestkrát v rozsahu přibližně šesti měsíců, vždy po dokončení příslušné lekce. Všechna testování zahrnovala vyplnění dvou čtecích testů v časovém rozsahu dvaceti minut. Při výběru jednotlivých testovacích technik jsem se snažila o to, aby byly použity techniky pro testování porozumění čtení co nejvhodnější. Další část výzkumu se týká obtížnosti výše zmíněných testovacích formátů měřených na základě správných odpovědí u každého z nich.

Ohledně testových formátů vhodných pro žáky devátých ročníků, tedy žáky pokročilé, domnívám se, že využití každého formátu má svá pozitiva, ale i negativa. Z výzkumu vyplývá, že podle výsledků je nejvhodnější technikou pro testování porozumění čtení typ testu s výběrem odpovědi a tím je ´multiple-choice´. Důvodem může být i to, že formát odpovědi není pro žáky matoucí, a že je hodnoceno pouze a jenom čtení. Nicméně, v neprospěch formátu hovoří fakt, že v některých případech test hodnotí spíše vyhledávání správné odpovědi v textu, spíše než porozumění textu samotnému. Také je nezbytné eliminovat žákovo případné hádání, čehož lze dosáhnout

několika způsoby, například upravením bodového systému. Přiřazovací úlohy jsou svou povahou velice podobné testům s výběrem odpovědi. V neprospěch tohoto testového formátu hovoří fakt, že je do jisté míry neautentický, jelikož se s ním žáci v reálných situacích velmi zřídkakdy setkají. Nebezpečí hrozí také v případě, pokud je v testu příliš položek a test se pro žáka stává spíše hádankou při hledání správných odpovědí, obzvláště při zařazení údajů navíc. Hodnocení techniky ´true/false´je do jisté míry velice podobné technice ´multiple-choice Jedním způsobem je například zadání třetí možnosti (´nezmíněno´) při psaní tohoto formátu testů. Při testování porozumění textu u žáků na pokročilé úrovni bychom za méně vhodnou mohli považovat také úlohu typu "picture-cued', jelikož žáci jsou vystaveni minimálnímu či dokonce žádnému čtení a informace získávají pouze z obrázku. V neprospěch testovací úlohy typu "cloze" mluví to, že slova musí být doplňována v závislosti na kontextu. Použití techniky „cloze elide" je rovněž sporné, jelikož ve své podstatě směřuje proti tomu, co je cílem rozvoje dovednosti čtení s porozuměním. Nevýhodou úloh s otevřeným koncem (short answer, summary) je jejich subjektivní skórování a v mnohých případech i fakt, že žáci jsou vystaveni psaní, což někdy vede k tomu, že učitel místo hodnocení čtení hodnotí psaní.

Při sestavování testů jsem se u každého z nich zaměřila nejprve na určení cíle. Co se týče typu testů, všechny testy jsou přímé, což testování čtení ve většině případů bývá. Dalším krokem bylo určit, na které čtecí mikro a makro-dovednosti se chci zaměřit. Jak jsem se již zmínila, při výběru testových formátů jsem se snažila o to, aby byly představeny formáty nejběžněji používané při testování čtení. Po konzultování odborné literatury jsem navrhla možné způsoby skórování u všech testů, z nichž všechny byly skórovány objektivně, což je pro testování čtení typické. S ohledem na validitu a reliabilitu testů, tato kritéria byla zajištěna konzultováním odborné literatury.

Závěrem bych chtěla dodat, že neexistuje nic takového jako nejlepší metoda, technika, testový formát, nebo nejlepší test. Test, který je ideální pro jeden účel může být zcela nevhodný pro jiný účel a formát testu, který může velice dobře fungovat v jedné situaci může být nevhodný v jiné situaci. Výběr testovací metody záleží na cílech, účelu a také důležitosti daného testu. Má diplomová práce se pokusila čtenářům poradit, jak postupovat při sestavování čtecích testů, ale jejím cílem bylo také ukázat, že čtení a zároveň i jeho testování jsou nedílnou součástí v procesu osvojování cizího jazyka.

BIBLIOGRAPHY:

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

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

Brown, Douglas. H. *Language Assessment. Principles and Classroom Practices.* New Jersey: Pearson Education, 2004.

Brown, Douglas H.. *Teaching by Principles.* New Jersey: Prentice Hall Regents, 1994

Duffy, Gerald G. Roehler, Laura R. *Improving Classroom Reading Instruction.* Caledonia: McGraw-Hill, 2004.

Gavora, Peter. *Úvod do pedagogického výzkumu.* Brno: Paido, 2000.

Greenwood, Jean. *Class Readers.* Oxford: Oxford University Press, 1988.

Harris, M. and P. McCann. *Assessment.* Oxford: Heinemann, 1994.

Heaton, John Brian. *Writing English Language Tests*. New York: Longman Inc, 1988.

Heaton, John Brian. *Classroom Testing*. London: Longman., 1990.

Hughes, Arthur. *Testing for Language Teachers*. Cambridge: Cambridge University Press, 1989.

Kalhous, Zdeněk. Obst, Otto. *Školní didaktika.* Praha: Portál, 2002.

Madsen, Harold. S. *Techniques in Testing.* Oxford: Oxford University Press, 1983.

Nuttal, Christine. *Teaching Reading Skills in a Foreign Language.* Oxford: Macmillan, 2005.

Simmons, John. Palmer, Barbara. *Reading by Doing*. Chicago: Lincolnwood, 1988.

Slavík, Jan. *Hodnocení v současné škole*. Praha: Portál, 1999.

Ur, Penny. *A Course in Language Teaching.* Cambridge: Cambridge University Press, 1996.

Walter, Catherine. *Authentic Reading*. Cambridge: Cambridge University Press, 1992.

Weir, Cyril J. *Communicative language testing.* Reading: Prentice Hall International, 1990.

Ytreberg, Lisbeth, Scott, Wendy. *Teaching English to Children*. London: Longman, 1991.

Internet sources:

Ballantyne, Christina. "Multiple-Choice Tests", last updated 30 October 2003 [viewed 18 April 2005]:

http://www.tlc.murdoch.edu.au/eddev/evaluation/mcq/mctests.html#strengths

Kitao, Kathleen. "Articles on Testing by Kathleen S. Kitao and Kenji Kitao." Testing Reading, last updated 4 April 1997 [viewed 14 March 2005]:

http://www.ilc2.doshisha.ac.jp/users/kkitao/library/article/test/reading.htm

"Constructing True/false Tests", last updated 15 June 1999 [viewed 18 April 2005]:

http://www.studygs.net/tsttak2.htm

"Preparing Multiple-Choice and Matching Tests", last updated 16 September 2000 [viewed 18 April 2005]:

http://www.nd.edu/~sharmati/reading.html

UW Teaching Academy. "Exam Question Types and Student Competencies. Matching Questions", last updated 2 June 2001 [viewed 18 April 2005]:

http://wiscinfo.doit.wisc.edu/teaching-academy/Assistance/course/matching.htm

UW Teaching Academy. "Exam Question Types and Student Competencies. True/false Questions", last updated 2 June 2001 [viewed 18 April 2005]:

http://wiscinfo.doit.wisc.edu/teaching-academy/Assistance/course/true.htm

UW Teaching Academy. "Exam Question Types and Student Competencies. Multiple-Choice Questions", last updated 2 June 2001 [viewed 18 April 2005]:

http://wiscinfo.doit.wisc.edu/teaching-academy/Assistance/course/mult.htm

UW Teaching Academy. "Exam Question Types and Student Competencies. Short-Answer Questions", last updated 2 June 2001 [viewed 18 April 2005]:

http://wiscinfo.doit.wisc.edu/teaching-academy/Assistance/course/sanswer.htm

APPENDIX 1: Reading test – Unit 2

Reading test – Unit 2[13]

1/ Read the original text and put the pictures in the correct order

The correct order of the pictures is: _____F, I, D, C, E, B, ?, ?, H_____

_____ S

2/ Read the original text and write down if the following sentences are T (true). F (false) or the text DS (doesn't say anything).

1/ You can make more than 12 cakes with these ingredients _____F_____

2/ You bake the cakes for longer than one minute _____T_____

3/ You put grated cheese under the potato cakes _____F T_____

4/ If you boil the potatoes for too long, they will not be nice _____F_____

5/ Before you put the cakes on a tray, you must take put them in the oven _____DS_____

6/ Butter and margarine go in one bowl together with flour and the mashed potatoes

_____F_____

7/ It is fine if the potato cakes are black on both sides _____F T_____

8/ You need eight ingredients for baking the potato cakes _____T_____

9/ You must slice the potatoes with a sharp knife _____DS_____

10/ You need less than half a kilo of potatoes for this recipe _____DS T_____

_____5 /10

Weighting is applied in this exercise in order to discourage guessing. It is done by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

6 – (4/3) = 5

Total score _____10 18

3

---

[13] See appendix 2 for the results, author of this test is in bold faced letters

APPENDIX 2: Reading test 1 – Unit 2

Reading test 1 – Unit 2

Student	Test 1	Exercise 1	Exercise 2	Grade
1	14	7	7	2
2	13	4	9	3
3				
4	**10**	**5**	**5**	**3**
5	10	4	6	3
6	17	7	10	1
7	18	8	10	1
8	17	8	9	1
9	11	5	6	3
10				
11	12	5	7	3
12	8	3	5	4
13				
14	18	8	10	1
15	12	4	8	3
16	18	8	10	1
17	15	6	9	2
18	16	8	8	2
19	9	3	6	4
Total 16	18	8	10	2,31

Index of difficulty:

Exercise 1 – ordering/picture-cued test

P = 100 x (ns/n)

P = 100 x (5/16) = **31,25**

Exercise 2 – true/false test

P = 100 x (ns/n)

P = 100 x (4/16) = **25**

Reading test – Unit 3[14]

1/ Read the original text and circle the correct answer.

1/ The most important meal in Britain is

A/ breakfast   B/ lunch   C/ dinner

2/ What meal do people eat after noon?

A/ breakfast   B/ lunch   C/ dinner

3/ What meal was quite big in the past but is not very today?

A/ breakfast   B/ lunch   C/ dinner

4/ Which meal can people get in canteens?

A/ breakfast   B/ lunch   C/ dinner

5/ Which meal do people also call supper?

A/ breakfast   B/ lunch   C/ dinner

6/ Which meal do people prepare food from other countries for?

A/ breakfast   B/ lunch   C/ dinner

7/ Which meal is quite big at weekends?

A/ breakfast   B/ lunch   C/ dinner

8/ Which meal has ham, cheese or vegetable between two slices of bread?

A/ breakfast   B/ lunch   C/ dinner

/8

2/ Read the original text and correct the false sentences. Cross out the wrong word and write the correct one. Some of the sentences are correct. If you find a correct sentence, write C (correct) next to it.

1/ Lunch is the most important meal of the day.

2/ British people often eat sauces in lunchtime.

3/ Many people do not eat meat.

4/ British people usually prepare big breakfast only on Saturdays.

[14] See appendix 4 for the results, author of this test is in bold faced letters

5/ There are toasts in a typical British lunch.

6/ Children can have their lunch at school or at home.

7/ A lot of people like to cook food from other countries.

8/ Grapes are peoples' favourite fruit.

9/ It is quite usual to have something sweet after dinner.

10/ Food from China or India is very popular nowadays.

11/ In some parts of Britain, 'dinner' means the same thing as 'breakfast'.

/11

Total score   /19

APPENDIX 4: Reading test 2 – Unit 3

Reading test 2 – Unit 3

Student	Test 2	Exercise 1	Exercise 2	Grade
1	15	8	7	2
2	13	7	6	3
3	**19**	**8**	**11**	**1**
4	15	8	7	2
5	12	5	7	3
6	15	6	9	2
7	19	8	11	1
8	17	8	9	1
9				
10	18	7	11	1
11	13	7	6	3
12	8	3	5	4
13	13	7	6	3
14	19	8	11	1
15	10	5	5	3
16				
17	15	8	7	2
18	15	7	8	2
19	7	3	4	5
Total 17	19	8	11	2,29

Index of difficulty:

Exercise 3 – multiple-choice test

P = 100 x (ns/n)

P = 100 x (8/17) = **47,05**

Exercise 4 – true/false test

P = 100 x (ns/n)

P = 100 x (4/17) = **23,52**

APPENDIX 5 – Reading test – Unit 4

Reading test – Unit 4[15]

1/ Read the original text and circle the correct answer

1/ Russell never sends Valentine cards

A/ because he is very lazy   B/ because he doesn't like any girl in his class

C/ because he is very popular with girls in his class

2/ Greg is in love with Nicola

A/ because she is very kind and pretty   B/ because she likes him too

C/ because she sits near him at class

3/ Greg bought the Valentine card

A/ because Peter liked it too   B/ because he thought that the words in

very nice   C/ because it was cheaper than other cards

4/ Greg wanted to take the card from Nicola's desk

A/ because he didn't put down his name   B/ because Amy saw him putting it

there   C/ because he heard that Nicola has a card for Russell

5/ Greg, Peter, Nicola, Amy, Russell and other children went out of the classroom

A/ because the school finished that day   B/ because they had free time between

two lessons   C/ because they went to the school canteen

6/ Greg was very happy when he came back to the classroom

A/ because he really liked what Nicola did to Russell   B/ because he talked to

Nicola outside   C/ because Nicola likes him too

·/ /6

Weighting is applied in this exercise in order to discourage guessing. It is done by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

6-(2/3) = 1

---

[15] See appendix 6 for the results, author of this test is in bold faced letters

2/ Read the original text, then read the sentences and replace the underlined pronouns with the nouns from the list. You can use some of them more than one time.

Russell   Greg   Nicola   cards   the card   Amy   Peter

1/ <u>They</u> left the building after lesson – *Greg, Peter*

2/ <u>She</u> really liked what Nicola did to Russell - *Amy*

3/ <u>It</u> cost more money than the rest – *Greg uri*

4/ <u>They</u> entered the classroom together – *Greg*

5/ <u>He</u> reminded him of the St. Valentine's Day - _____

6/ Everyone made fun of <u>him</u> – *Greg*

7/ <u>He</u> wasn't happy at all - *Peter*

8/ I will receive a great number of <u>them</u> – _____

9/ <u>He</u> was very popular with girls – *Russel*

4/9

Total score   *5* /15

79

APPENDIX 6: Reading test 3 – Unit 4

Reading test 3 – Unit 4

Student	Test 3	Exercise 1	Exercise 2	Grade
1	11	6	5	2
2	10	4	6	3
3	14	6	8	1
4	11	6	7	2
5				
6	13	4	9	2
7	14	5	9	1
8	13	6	7	2
9	8	3	5	3
10	14	6	8	1
11	9	4	5	3
12	**5**	**1**	**4**	**5**
13	10	5	5	3
14	15	6	9	1
15	8	4	4	3
16	13	4	9	2
17	11	4	7	2
18				
19	6	3	3	4
Total 17	15	6	9	2,38

Index of difficulty:

Exercise 5 – multiple-choice test

P = 100 x (ns/n)

P = 100x (6/17) = **35,29**

Exercise 6 – matching test

P = 100 x (ns/n)

P = 100x (4/17) = **23,52**

APPENDIX 7 – Reading test – Unit 5

Reading test – Unit 5[16]

1/Read the original text and match the sentences. Write the letters in the gaps provided. You do not need to use all the answers.

1/ It is located in California
2/ It is the year when USA got its independence
3/ It is the population of the country
4/ It is a building where the US president lives
5/ It is the third longest in the world
6/ It is on the south boarder of the USA
7/It is a little state in the east of USA
8/ It is the biggest city in the country
9/ It is the fourth largest in the world
10/ It is on the northern boarder of the USA

A/ the United States of America
B/ the Mississippi-Missouri River
C/ New York
D/ 8 million
E/ Canada
F/ Chicago
G/ Los Angeles
H/ 260 million
I/ Washington D.C.
J/ Mexico
K/ the White House
L/ California
M/ 1776
N/ Florida

---

[16] See appendix 8 for the results, author of this test is in bold faced letters

O/ Delaware

2/ Look at the map and match the following descriptions and the geographical names from the map. Write them in the gaps provided.

1/ The American state on the west coast of the USA - _____

2/ The country bordering with the USA in the north - _____

3/ One of the small American states above New York - _____

4/ The city on the coast of the Great Lakes - _____

5/ The most southern state of the USA - _____

6/ The country bordering with the USA in the south - _____

7/ The longest river in the USA - _____

8/ The state between two cities - _____

9/ The city under Delaware - _____

10/ The state on the south east coast of the USA - _____

12/ The biggest city in the west coast of the USA - _____

12/12

Total score   *22* /22

APPENDIX 8: Reading test 4 – Unit 5

Reading test 4 – Unit 5

Student	Test 4	Exercise 1	Exercise 2	Grade
1				
2	**22**	**10**	**12**	**1**
3	21	9	12	1
4	15	8	7	3
5	12	6	6	3
6				
7	21	10	11	1
8	20	8	12	1
9	11	4	7	4
10	20	10	10	1
11	12	6	6	3
12	10	6	4	4
13	14	8	6	3
14	22	10	12	1
15				
16	19	10	9	1
17	13	6	7	3
18				
19	9	4	5	4
Total 15	22	10	12	2,24

Index of difficulty:

Exercise 7 – matching test

P = 100 x (ns/n)

P = 100 x (5/15) = **33,33**

Exercise 8 – picture-cued test

P = 100 x (ns/n)

P = 100 x (4/15) = **26,26**

APPENDIX 9 – Reading test 5 – Unit 6

Reading test – Unit 6[17]

1/ Fill the sentences (A-G). Write the letter in the gaps provided. There are three extra sentences that you do not need.

Luke Ryan opened his eyes slowly. He was on a bed in a strange room. Kirsty was there, too. "Are you all right?" she asked quietly. "It doesn't matter. You're both going to die." It was Luke's old enemy, Doctor Dread. He laughed horribly. "I've got the computer chip now. So goodbye, Mr Ryan." _____

"It looks like the end," said Kirsty sadly. "No, we must stop him," said Luke. "With that chip he can destroy all the world's satellites." "I know," said Kirsty. "That guard is very slow. I think we can escape easily." She pointed to the blanket on the bed and whispered to Luke. "That's a good idea," said Luke and he started to shout very loudly. "Oh, my stomach. Help! My stomach! Help!"

_____ "What's all that noise?" he said. Kirsty threw the blanket over the guard's head. Luke hit him hard in the stomach and quickly grabbed his gun and the keys. "You did that very well," said Luke, as they locked the guard in the room and ran out. "Oh, it was easy," said Kirsty.

_____ Luke pushed the driver out and they drove as fast as possible to the airport. When they arrived at the airport, they saw a small plane. "That's Dread's plane," said Luke. They ran to the plane. It was empty, so they got in and hid behind the seats. Soon Doctor Dread got into the plane. _____

_____ "With that chip I'm going to be master of the world," he said proudly. "No, you aren't," shouted Luke. He jumped up from behind the seat and grabbed the bag. "Give that to me," said Doctor Dread. He quickly grabbed the bag and pulled hard. But he pulled too hard. The door behind him suddenly opened and he fell out. With the bag in his hand he fell down and down and down into the sea.

_____

---

[17] See appendix 8 for the results, author of this test is in bold faced letters

A/ There was a car in the street.

B/ Luke and Kirsty opened the door.

C/ The car was very small.

D/ Kirsty flew the plane back to the airport and they landed safely.

E/ He left the room and the guard locked the door.

F/ The guard rushed in.

G/ He carefully put a bag on the seat next to him, started the plane and took off.

H/ Ten minutes later, when they were over the sea, he looked at the bag.

7 /6

2/ Look at the pictures in the original text and match them with these sentences. There is one sentence that you do not need to use.

1/ The guard looked in the room through the little door window _____

2/ Luke grabbed the bag _____

3/ Luke drove the car to the airport _____

4/ Luke and Kirsty threw a blanket over the guard's head _____

5/ The door opened and Dr. Dread fell out of the plane _____

6/ Dr. Dread visited Luke and Kirsty in their room _____

7/ Kirsty had an idea and she told Luke about it _____

8/ Kirsty and Luke got in the plane _____

9/ Luke and Kirsty watched Dr. Dread as he fell out of the plane _____

10/ Kirsty drove the car to the airport _____

6 /9

10

Total score 9 /15

85

APPENDIX 10: Reading test 5 – Unit 6

Reading test 5 – Unit 6

Student	Test 5	Exercise 1	Exercise 2	Grade
1	13	6	7	2
2	11	4	7	3
3	15	6	9	1
4	11	4	7	3
5				
6	12	6	6	2
7				
8	13	6	7	2
9	7	2	5	4
10	15	6	9	1
11	7	3	4	4
12	6	2	4	4
13	9	4	5	3
14	15	6	9	1
15	10	4	6	3
16				
17	**10**	**4**	**6**	**3**
18				
19	6	2	4	4
Total 15	15	6	9	2,66

Index of difficulty:

Exercise 9 – matching test

P = 100 x (ns/n)

P = 100 x (6/15) = **40**

Exercise 10 – picture-cued test

P = 100 x (ns/n)

P = 100 x (3/15) **= 20**

Reading test – Unit 7[18]

1/ Read the original text and choose the correct answer

1/ The type of entertainment that people like the most in Britain is.....

A/ cinema   B/ going to the library   C/ TV

2/ Usual channels in Britain put on ....programmes

A/ various   B/ same   C/ similar

3/ The channels which are not free put on .... of programmes

A/ a similar type   B/ various types   C/ the same type

4/ Videos are favourite with people because..........

A/ are cheaper than cinema tickets   B/ people like sitting at their homes not in cinemas

C/ a video shop is in almost every town but cinemas only in large towns

5/ When you go to the theatre, you can see ....

A/ people whose job is acting   B/ people whose hobby is acting   C/ both types of people

6/ If you see a pantomime in the theatre, it will be................

A/ very sad   B/ very funny   C/dramatic

7/ If they lend you books in a public library, you ....

A/ should pay   B/ don't have to pay   C/ must pay

/7

Weighting is applied in this exercise in order to discourage guessing. It is done by dividing the number of wrong answers by the number of distractors and subtracting this number from the number of correct answers.

7- (1/6) = 7

---

[18] See appendix 8 for the results, author of this test is in bold faced letters

2/ Read the original text and correct the false sentences. Cross out the wrong word and write the correct one. Some of the sentences are correct. If you find a correct sentence, write C (correct) next to it.

1/ Television is the least favourite type of entertainment in the UK. _____

2/ They have a theatre in every large town. _____

3/ There is a private library in all towns and a lot of villages. _____

4/ Most theatres show a comedy in November and December. _____

5/ Digital and satellite channels are not free in the UK. _____

6/ Children can watch children shows from four o'clock to six thirty. _____

7/ People can buy videos in video shops. _____

8/ If you want to borrow something to read in a library you don't have to pay.

_____

5 /8

Total score   /15

APPENDIX 12: Reading test 6 – Unit 7

Reading test 6 – Unit 7

Student	Test 6	Exercise 1	Exercise 2	Grade
1	14	7	7	1
2				
3	15	7	8	1
4	11	5	6	3
5	7	3	4	4
6	12	6	6	2
7	15	7	8	1
8	15	7	8	1
9				
10	14	7	7	2
11	9	4	5	3
12	5	2	3	5
13	8	3	5	4
14				
15	**11**	**6**	**5**	**3**
16				
17	10	4	6	3
18				
19	5	2	3	5
Total 14	15	7	8	2,78

Index of difficulty:

Exercise 11 – multiple-choice test

P = 100 x (ns/n)

P = 100 x (5/14) = **35,71**

Exercise 12 – true/false test

P = 100 x (ns/n)

P = 100 x (3/14) = **21,42**

APPENDIX 13

Index of difficulty:

Test format Exercise	Multiple-choice	Matching	True/false	Picture-cued
1				31,25
2			25	
3	47,05			
4			23,25	
5	35,29			
6		23,52		
7		33,33		
8				26,26
9		40		
10				20
11	35,71			
12			21,42	
Total	**39,35**	**32,28**	**27,28**	**25,84**

APPENDIX 14

Grading system

Grade	Number of per cent
1	89 – 100
2	73 – 88
3	55 -72
4	39 – 54
5	0 – 38

D SAM'S POTATO CAKES

Comprehension

1 Sam is making potato cakes.

a Look at the pictures. What order do you think they go in?

b 🔊 51 Read and listen to the recipe. Number the pictures in the correct order.

> We write: 400g potatoes
> We say: four hundred grams **of** potatoes

> I'm going to make some potato cakes. They're delicious. For this recipe you need a few potatoes, some flour, some butter or margarine, some grated cheese, a bit of salt and a bit of oil.

Potato cakes

Ingredients

You will need:

400g	potatoes
125g	flour
40g	butter or margarine
40g	grated cheese
	salt
	oil
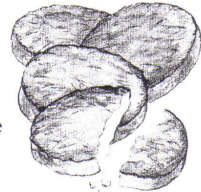
Instructions

1 Peel the potatoes and slice them.

2 Put the potatoes in a saucepan with some water and a bit of salt. Boil the potatoes for 20 minutes.

3 Pour away the water and mash the potatoes in the saucepan.

4 Put the mashed potatoes, the flour and the butter or margarine into a bowl. Mix them with a fork.

5 Take the mixture out of the bowl and roll it. Make 12 potato cakes like this:

⬭ ⥮ 1 cm

6 Put the oil in a frying pan. Fry the potato cakes until they are brown on both sides.

7 Take the cakes out of the frying pan and put them on a tray. Cover the cakes with the grated cheese.

8 Put the tray into the oven for two minutes.

MEALS AND MEAL TIMES

92

Read about eating in Britain.
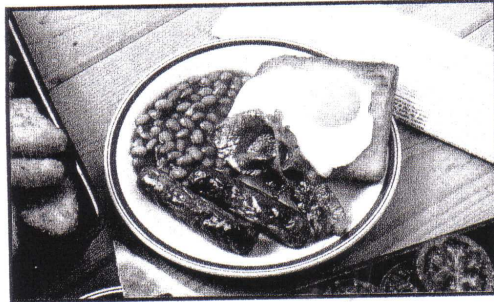
The first meal of the day is breakfast. The traditional British breakfast is very big: bacon, eggs, sausages, tomatoes, mushrooms, baked beans and fried bread! But nowadays people don't usually cook a big breakfast every morning – they only have a cooked breakfast on Sundays, or when they stay in a hotel.

Most people have a light breakfast. These are some of the typical things that people have for breakfast:

- toast with jam, honey or marmalade
- cereal with milk and sugar
- fruit (Grapefruit is very popular.)
- yoghurt
- orange juice
- coffee or tea

Between 12 o'clock and two o'clock people have lunch. This is often a light meal of soup, sandwiches and/or a salad. Most children have their lunch at school. They can take sandwiches from home or they can buy food at the school cafeteria.

The main meal of the day is dinner. People eat this between about six and seven o'clock in the evening. A traditional dinner is meat or fish with potatoes and vegetables, but many people nowadays cook foreign food, and lots of people are vegetarian, as well.
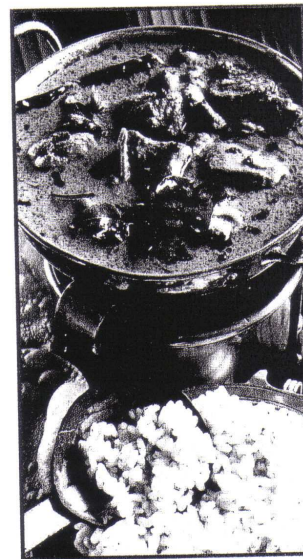
Pasta and pizzas are very popular, and Indian food, such as curry with rice. Chinese food is also a favourite. There are lots of cookery programmes on TV and people like to try new dishes. After the main course some people have dessert (also called 'pudding') – such as fruit, chocolate cake, or apple pie with ice-cream or custard.

In some areas of Britain, 'lunch' is called 'dinner', and 'dinner' is called 'tea' or 'supper'.

Comprehension

1 🎞 67 Look at the two cards. Read and listen to the story.

a Who is each card from?

b Who is it to?

c What does it say inside the second card?

2 Read the story again. Who or what do these words describe?

beautiful	happy	proud
gorgeous	tall	nice
miserable	red	big
expensive	white	good-looking
intelligent		

The Valentine's Card

'It's 14 February tomorrow – Valentine's Day,' said Peter. 'Are you going to send any cards?'

'I don't know,' said Greg, but it wasn't true.

'Well, I never send any Valentine's cards, but I'm going to get lots of them.' It was Russell Jenkins, and he was probably right. He was the tallest and best-looking boy in the class. All the girls liked him. Unfortunately, he knew it.

After school Greg went to the shop to buy a card for Nicola. She sat near Greg in class and he thought she was the most gorgeous girl in the world. She had beautiful eyes and she was nice, too. She laughed a lot and she was very intelligent. Greg chose a card. It said: 'My love for you is as high as a mountain and as deep as the sea. Be my Valentine.' It was more expensive than the other cards, but Greg liked it.

The next day Greg went into the classroom early and put the card in Nicola's desk. Nicola came in with her friend, Amy.

'Are you really going to give that card to Russell?' Amy asked.

'Yes, I'm going to give it to him at break,' said Nicola and they both laughed. Greg felt miserable. He wanted to get his card from Nicola's desk, but it was too late. Fortunately, his name wasn't in the card.

Nicola found the card, opened it, smiled and looked round. Greg didn't look at her, but he knew that his face was as red as a beetroot.

Break came and Greg went outside with Peter. Russell was there. He looked as proud as a peacock with all his cards.

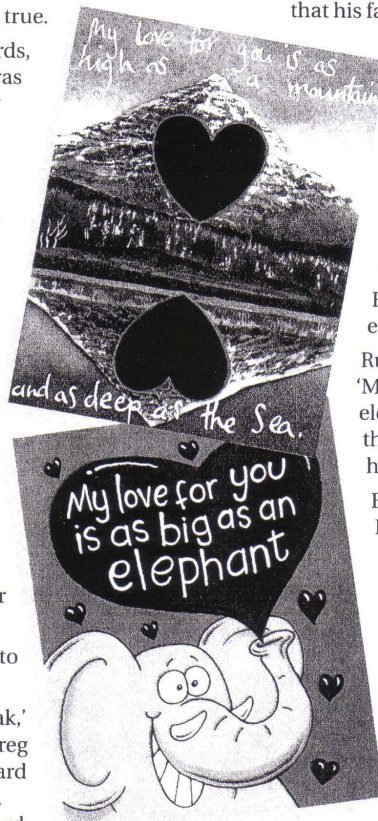'You see,' he said. 'I didn't send any cards, but I got lots of them.'

'Here's another card for you, Russell.' It was Nicola. Greg felt even more miserable.

Russell opened the envelope. 'My love for you is as big as an elephant,' he read. He opened the card. 'But then, so is your head!'

Everyone laughed. 'Big head! Big head!' they shouted.

'Nice one, Nicola,' said Peter. Greg felt a lot happier.

Back in the classroom Greg sat down and opened his desk. Then he saw the most amazing thing – a white envelope. He looked round and the most gorgeous girl in the world smiled at him.
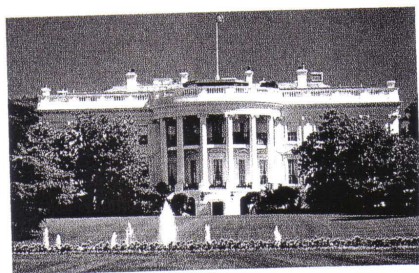
THE USA 🔊 93

Now here's something about the USA.

The United States of America is the fourth largest country in the world. It is bigger than the whole of Europe. There are five time zones in the USA. So, when it's twelve o'clock in New York, it's only eight o'clock in Alaska. The third longest river in the world is in the USA. It's the Mississippi–Missouri River.
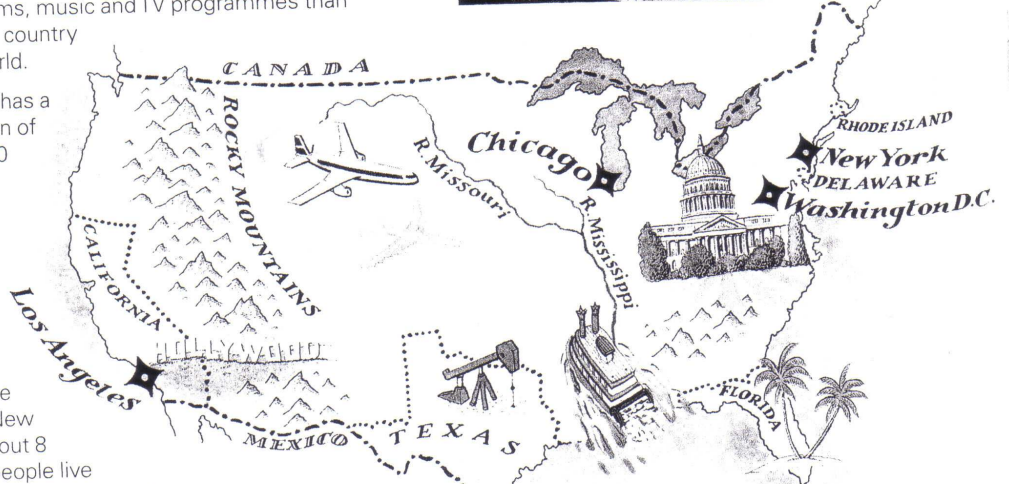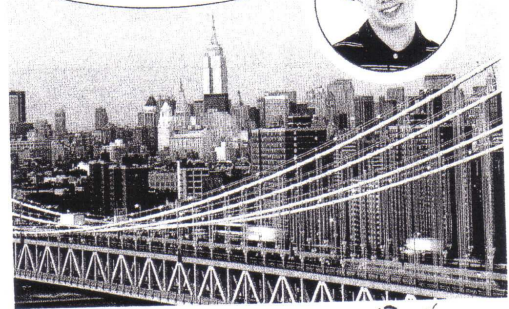
The USA is the richest country in the world. It produces more food, iron, cars, aeroplanes, books, films, music and TV programmes than any other country in the world.

The USA has a population of about 260 million people. Most of them live in cities. The largest city in the USA is New York. About 8 million people live there. The second largest city is Los Angeles in California and the third largest is Chicago.

There are fifty states in the United States. Some of the eastern states, like Rhode Island and Delaware, are quite small, but others, like Texas, California and Alaska, are bigger than a lot of countries. Each state has got its own government and its own capital city. There is a federal capital for the whole of the USA in Washington DC. The US President lives there, in the White House.

This is the flag of the USA. It's called 'The Stars and Stripes'. There is one star for each of the fifty states. The thirteen stripes are for the original thirteen states that formed the USA in 1776, when they got their independence from Britain.

The first settlers in the USA came from Britain, but now the USA has people from all over the world – from Europe, Asia, Africa and South America.

Comprehension

1 a Look at the pictures. Find the names of the people in the story.

b What is happening in each picture?

2 📼 75 Read and listen to the story. Number the pictures in the correct order.

Dr. Dread

Luke Ryan opened his eyes slowly. He was on a bed in a strange room. Kirsty was there, too.

'Are you all right?' she asked quietly.

'It doesn't matter. You're both going to die.' It was Luke's old enemy, Doctor Dread. He laughed horribly. 'I've got the computer chip now. So goodbye, Mr Ryan.' He left the room and the guard locked the door.

'It looks like the end,' said Kirsty sadly.

'No, we must stop him,' said Luke. 'With that chip he can destroy all the world's satellites.'

'I know,' said Kirsty. ' That guard is very slow. I think we can escape easily.' She pointed to the blanket on the bed and whispered to Luke.

'That's a good idea,' said Luke and he started to shout very loudly. 'Oh, my stomach. Help! My stomach! Help!'

The guard rushed in. 'What's all that noise?' he said. Kirsty threw the blanket over the guard's head. Luke hit him hard in the stomach and quickly grabbed his gun and the keys.

'You did that very well,' said Luke, as they locked the guard in the room and ran out.

'Oh, it was easy,' said Kirsty.

There was a car in the street. Luke pushed the driver out and they drove as fast as possible to the airport. When they arrived at the airport, they saw a small plane.
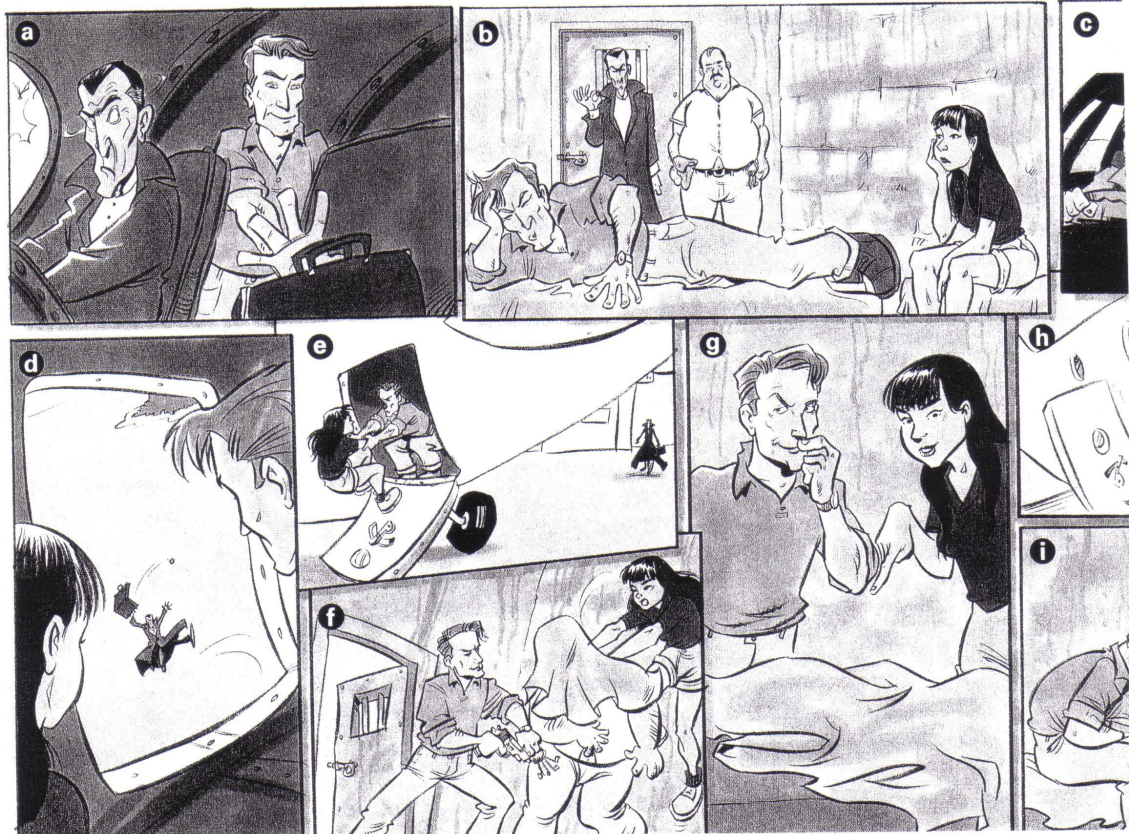
'That's Dread's plane,' said Luke. They ran to the plane. It was empty, so they got in and hid behind the seats.

Soon Doctor Dread got into the plane. He carefully put a bag on the seat next to him, started the plane and took off. Ten minutes later, when they were over the sea, he looked at the bag.

'With that chip I'm going to be master of the world,' he said proudly.

'No, you aren't,' shouted Luke. He jumped up from behind the seat and grabbed the bag.

'Give that to me,' said Doctor Dread. He quickly grabbed the bag and pulled hard. But he pulled too hard. The door behind him suddenly opened and he fell out. With the bag in his hand he fell down and down and down into the sea.

Kirsty flew the plane back to the airport and they landed safely.

Grammar

3 a Complete the table.

Adjectives and adverbs

Adjectives describe nouns or pronouns.
That guard is very ....
'Oh, it was ....', said Kirsty.

Adverbs describe verbs.
Luke Ryan opened his eyes ....
I think we can escape ....

b What letters do adverbs end with? Find more adverbs in the story.

Be careful with these!

c Complete these sentences. Are the missing words adverbs or adjectives?

1 You did that very ....
2 They drove as .... as possible to the airport.
3 He quickly grabbed the bag and pulled ....

4 Choose the correct word to complete the sentences.

1 The car was very *slow/slowly*.
2 Kirsty flew the plane *safe/safely*.
3 Luke can drive very *good/well*.
4 Doctor Dread grabbed the bag *quick/quickly*.
5 Luke made a *loud/loudly* noise.
6 Doctor Dread spoke *quiet/quietly*.
7 He was a *horrible/horribly* man.
8 'We're *safe/safely* now,' said Luke *happy/happily*.
9 Doctor Dread put the chip in the bag *careful/carefully*.
10 He pushed Luke very *hard/hardly*.

Writing

5 a You only have the last part of the story of Doctor Dread here. Think about what happened before this.
What happened to Luke?
Why were he and Kirsty in the room?
How did Doctor Dread get the computer chip?
Why did he want it?

b Work with three friends. Write the first part of the story.

c Act your story.

ENTERTAINMENT 📼 94

The most popular form of entertainment in Britain is television. There are five normal channels: BBC1, BBC2, ITV, Channel 4 and Channel 5. Each channel shows a wide range of different programmes. There are special programmes for children from four o'clock to six o'clock.

There are also several digital and satellite channels. You have to pay to watch these. Each channel usually shows just one kind of programme – cartoons, comedy programmes, sport, films, pop music and so on.

Here's some information about entertainment in Britain.

You only find cinemas in larger towns. Each cinema normally has two or more screens, so that they can show more than one film at a time.

Cinemas are quite expensive, so videos are very popular. You can hire videos from a video shop. There's a video shop in almost every town.
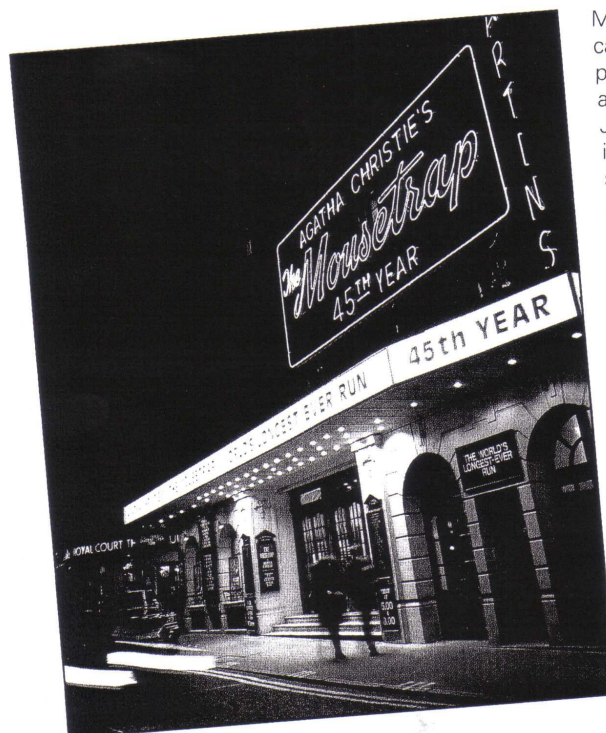
Most larger towns have got a theatre, where you can see plays and concerts. Some plays are by professional actors, but there are also a lot of amateur drama groups. In December and January most theatres put on a pantomime. This is a musical comedy based on a traditional story, such as *Cinderella* or *Sleeping Beauty*. The big cities like London, Glasgow and Birmingham have got several theatres.

All towns and a lot of villages have a public library, where you can borrow books. At a lot of libraries you can also borrow videos, CDs and CD-ROMs. The books are free, but you usually have to pay for the other things.

ÚDAJE PRO KNIHOVNICKOU DATABÁZI

Název práce	Testing Reading in English Language Teaching
Autor práce	Veronika Benešová
Obor	Učitelství anglického jazyka
Rok obhajoby	2006
Vedoucí práce	Mgr. Pavel Brebera
Anotace	Schopnost porozumět psanému textu je nezbytná pro úspěšnou komunikaci, proto je cílem výuky cizích jazyků osvojit si tuto dovednost a také ji testovat, aby žáci mohli adekvátně reagovat na informace získané z textů.
Klíčová slova	Testování a výuka/Testing and teaching Typy testů/Types of tests Validita, reliabilita/Validity, reliability Čtení/Reading Fáze při sestavování testů čtení/Stages of reading test construction Testovací techniky/Techniques of testing