

Západočeská univerzita v Plzni

**Fakulta pedagogická
Katedra anglického jazyka**

**Diplomová práce
PROBLÉMY DESIGNU MATERIÁLŮ PRO
TESTOVÁNÍ JAZYKŮ**

Miroslava Voláková

Plzeň 2013

University of West Bohemia

**Faculty of Education
Department of English**

Thesis
**DESIGN ISSUES IN LANGUAGE TESTING
MATERIALS**

Miroslava Voláková

Plzeň 2013

Tato stránka bude ve svázané práci Váš původní formulář *Zadání dipl. práce*
(k vyzvednutí u sekretářky KAN)

Prohlašuji, že jsem práci vypracoval/a samostatně s použitím uvedené literatury a zdrojů informací.

V Plzni dne 28. června 2013

.....
Miroslava Voláková

ACKNOWLEDGMENTS

I would like to express my thanks to my supervisor Mgr. Gabriela Klečková, PhD. for her guidance, support, time, patience, and suggestions. I would also like to thank to the participants of my research for their time and goodwill.

ABSTRACT

Voláková, Miroslava. University of West Bohemia. June, 2013. Design Issues in Language Testing Materials. Supervisor: Mgr. Gabriela Klečková, PhD.

The thesis deals with the possible impact of the visual design of a language test on students' perception of such a test. It provides information about the essential design rules and laws, and furthermore analyses their use in a didactic test in English created for the state school-leaving exam. The aim of the research carried out by means of usability testing was to find out design issues and suggest their possible solution.

TABLE OF CONTENTS

I. INTRODUCTION	1
II. THEORETICAL BACKGROUND	3
Design Principles	3
C. R. A. P. Rules	3
Gestalt Principles and Laws	5
Visual analysis of documents	8
Intra-level Design	8
Inter-level Design	9
Extra-level Design	9
Supra-level Designs	10
Technical Features of Language Testing	10
Validity	10
Reliability	14
Usability testing	17
Usability	17
Usability Testing	18
Limitations of usability testing	19
The process of usability testing	19
III. METHODS	22
Test Plan	22
Tested product	22
Purpose, goals, objectives	23

Research questions.....	24
Characteristics of participants.....	24
Testing method.....	25
Testing environment	25
Testing equipment.....	26
IV. RESULTS AND COMMENTARIES.....	27
Question #1	27
Question #2	28
Question #3	30
Question #4	32
Question #5	33
Question #6	35
Conclusion	37
V. IMPLICATIONS.....	39
Implications for Teaching.....	39
Limitations of the Research	41
Suggestions for further research	42
VI. CONCLUSION.....	43
REFERENCES	44
APPENDICES	46

LIST OF GRAPHS

<i>Graph 1.</i> Final times for task #3	31
<i>Graph 2.</i> Final times for task #4	32
<i>Graph 3.</i> Final times for task #5	34

I. INTRODUCTION

Many of us have come in touch with at least one test in our lives. In nowadays society, thriving to understand as many cultures as possible, it is highly probable we encountered language tests, too. These language tests might have been different kinds of tests ranging from those being carried out before being admitted to a school or a course through those encountered during the school years to those passed in order to get a language certificate. There are tens and hundreds of language tests being carried out every year.

If we asked students about the language tests they had taken during their education years, they might remember one or two, probably the most difficult ones, or the ones that made them the most proud of themselves. They might even recall the kinds of tasks the test consisted of. However, if we asked them whether the test was well designed, we would probably get a strange look from them.

When considering the topic of language testing and language tests, the areas that are discussed more often than any other are the technical features of tests - their validity, reliability, and also their practicality. Does the test measure what it is supposed to measure? Do the results correspond to the students' abilities? Is the test somewhat easy to grade? However, there is one area that we think is of the same importance, and which is often disregarded. It is the area of the visual design of the test, which is the focus of this thesis

In this work, I present an analysis of the design of one of the most important language tests in the Czech Republic nowadays - the state school-leaving exam in English, and examine how different design principles are applied, and whether the test contains parts that might make it difficult for students to work with, or might even affect the final result of the test.

The thesis consists of several logically built parts. First, it provides a theoretical background presenting some of the most essential rules and laws of design. Second, it introduces the methods used in conducting the research. Third, it presents the results obtained during the testing, and provides commentaries and explanations. Last but not least, it provides implications stemming from the research and its results.

II. THEORETICAL BACKGROUND

In the theoretical part, background information about the topic of the research can be found. This part is divided into sections introducing the basic rules of design and how they work, the visual analysis of the document, as well as some of the technical aspects of language testing. It also introduces the research method on a theoretical level.

Design Principles

C. R. A. P. Rules

The C. R. A. P. rules are one of the very basic principles of design. The abbreviation stands for contrast, repetition, alignment, and proximity. These are the most important components of design, as they help the overall structure be well coordinated, easier to understand, and easier to navigate. Even though people usually use these rules naturally and without giving it much thought, it is vital to understand how these work to be able to make the document express the desired information in a way that was intended for it.

Contrast. Contrast is one of the most effective ways to add visual interest to a page. It helps avoid elements that are merely similar by making them really different (Williams, 2008). Without contrast, all visual elements would look the same, monotonous (Landa, 2011). Contrast adds shape, form and dynamism to a design, and is even able to create a dramatic tension (Ambrose & Harris, 2007). It creates visual diversity, and makes difference between the elements by creating visual hierarchy of information (Landa, 2011). Contrast not only helps to distinguish elements from each other, but also makes it easier for readers to instantly understand the way the information is organized within a page or even a more complex structure. For the

contrast to be effective, it has to be strong enough, so that the reader is able to distinguish between the elements (Williams, 2008).

Repetition. In general, repetition helps the organization and strengthens the unity of a document. Designers often use repetition (e.g. using headings of the same height and weight) to make documents more consistent, to make the pages look like they actually belong together. Once the reader is familiar with the image or message of a certain item, they are likely to make an automatic connection when they come across it again (Ambrose & Harris, 2007). However, it is advised not to overuse repetition too much, as it might become annoying for the readers (Williams, 2008).

Alignment. Alignment stands for the placement of elements within a page, such as lining up the edges along common rows or columns (Lidwell, Høleden, & Butler, 2010). Alignment helps the elements have their place on a page. Nothing within a given document should be placed arbitrarily; every element should have a visual connection with another element. Aligned items create a stronger, more cohesive unit, thus they are easier to understand and categorize. Alignment adds certain stability and equilibrium to documents by making them well balanced, and thus improves the overall aesthetics of the document. Alignment can actually become a powerful means of leading a person through a design (Lidwell et al., 2010).

In the western world, designers usually choose to align bodies of text to the left, as it is a direction of reading people are used to. Center-aligned text blocks appear more ambiguous, and thus the page should always be designed so that readers could move in their normal moving pattern (i.e. left to right) (Lidwell et al., 2010; Weinschenk, 2011). Similarly to repetition, alignment should not be overused; there should never be more than one text alignment on a page (Williams, 2008).

Proximity. The rule of proximity says that items which relate to each other should be clearly grouped together. Several items in proximity to each other become one visual unit, which helps organize the information, reduces clutter, and thus gives the document a clear structure. This rule is linked with that of grouping, one of the gestalt principles, discussed further on in this work.

It is important for the reader to get as much information about the document as possible at first glance. Proximity helps clearly distinguish how many units there are within one page (e.g. units divided by headings), clearly identify the start and the finish of a document, and also organize the white space in a better way. There should be no more than 3-5 units per page, as more of them could create a clutter (Williams, 2008).

Gestalt Principles and Laws

Gestalt principles. Gestalt principles and laws are a set of perceptual rules, based on the German Gestalt School of Psychology founded in 1912, whose main representatives were Max Wertheimer, Kurt Koffka, Wolfgang Kohler, and later on also Rudolf Arnheim. Gestalt rules basically introduce the way people perceive. In general, Gestalt principles are engaged to increase the unity and consistency of a document (Hampe & Konsorski-Lang, 2010; Landa, 2011; Ware, 2012).

The whole vs. the sum of its parts. The very basic principle of Gestalt says, that in perception, the whole is larger than the sum of its parts. For example, when reading, the reader perceives each word first as a complete unit rather than seeing the individual letters (Hampe & Konsorski-Lang, 2010).

Figure and ground relationship. Another well-known principle of perception is the figure and ground relationship. It says that the form of an object is not more important than the form of the space around the object; the figure (i.e. an object) is

always seen in relation to the ground (i.e. the space surrounding it) (Lupton & Phillips, 2008). Both figure and ground have certain characteristics, which help to distinguish between them. Figure has a definite shape while ground is shapeless. Ground continues behind the figure. Figure seems to be closer with a clear location in space (Lidwell et al., 2010). Simply put, figure is something object-like, something perceived as a foreground while ground is what lies behind the figure (Ware, 2012).

When designing a document, we should seek a stable relationship between these two elements; figure and ground should always be clearly differentiated, as it makes the document clearer for the reader. Basically, there exist three types of the relationship between the figure and the ground:

Stable relationship. This relationship is the one designers usually aim for. In a stable relationship, the figure stands clearly from the ground.

Reversible relationship. This relationship appears in a document, when both the figure and the ground attract the attention of the reader equally and alternatively, coming out and receding.

Ambiguous relationship. Ambiguous relationship rises when the viewer is not able to find a focal point, as there is no discernable assignment of dominance in the document. The ambiguity of figure and ground can shift the result and impact of a document and the reader can interpret it in a different way than intended. Thus one of the essential skills of every designer is to be able to evaluate the tension between the figure and the ground (Lidwell et al., 2010; Lupton & Phillips, 2008).

Gestalt laws. Gestalt laws explain how people perceive, and in this way help designers place the elements within a page or a whole document. Authors slightly differ in the number of laws presented, however, the vast majority of them includes the five basic ones:

Law of proximity. Objects standing close to each other are perceived as grouped together. Texts belonging together should be grouped nearby (e.g. headlines should stand closer to the text that follows rather than the text preceding them) (Hampe & Konsorski-Lang, 2010).

Law of similarity (grouping). Objects similar in characteristics (e.g. a form, colour, size, or brightness) tend to be perceived as a group. Thus elements of bulleted lists, highlighted words, boxes, and other elements should be used consistently within a document. This applies also for underlining, boldface, colour, font size of different parts of text, symbols and icons (Hampe & Konsorski-Lang, 2010; Ware, 2012).

Law of closure. People perceptually tend to complete objects that have gaps in them or are not complete. Put in other words, open curves tend to be perceived as complete forms, because our mind has a tendency to produce a complete form, unit, or pattern. That is why we perceive tables, columns, boxes, and other elements as entities, because of their closed form, even if they are not complete or are broken by another element (Hampe & Konsorski-Lang, 2010; Steinfeld & Maisel, 2012; Landa, 2011).

Law of symmetry. Symmetrical shapes and forms are perceived as forming a group, even in spite of distance (Hampe & Konsorski-Lang, 2010). Symmetrical arrangements tend to stand out from the background; symmetry enhances perception and helps people to remember relationships – symmetrical organization is easier to remember. On the other hand, items out of place in an otherwise symmetrical arrangement will stand out more easily and thus the reader will notice them more (Steinfeld & Maisel, 2012).

Law of continuity. People tend to see continuous visual elements as visual entities rather than ones making abrupt turns. A group of similar objects is perceived

as a line in the smoothest path. E.g. a bulleted list will thus be perceived as a line, like a string of beads (Hampe & Konsorski-Lang, 2010; Steinfeld & Maisel, 2012).

Visual analysis of documents

One way of looking at the visual vocabulary of documents is to distinguish between the different levels of design from local to large-scale. Kostelnick & Roberts (1998) recognize four basic levels of design: intra, inter, extra, and supra. The first two aforementioned levels pertain primarily to text design; extra level pertains primarily to non-textual elements (e.g. data displays, pictures etc.); and supra level refers to the large-scale design of the whole document. Furthermore, each of these levels may contain design elements in three coding modes: textual, spatial, graphic. They supply the raw materials of design like the words, numbers, and graphic elements (e.g. lines, textures, shading etc.), and the spatial positioning of these elements on a page. Together, the levels of design and their modes create the visual matrix of a document (Kostelnick & Roberts, 1998).

Intra-level Design

Intra-level design consists of linear components. It controls local variations of text and creates the atoms and particles of the visible text. Individually, intra-level effects are small, but they are multiplied many times throughout, and thus have a huge effect on the visual language of a document.

The textual mode of intra-level design consists of the typeface selection, type size, and the treatment of the typeface (i.e. whether it is in italics, bold, roman, upper or lower case etc.). The spatial mode governs the flow of letters and words in a line of text, and consists mostly of local spacing between textual units. The graphic mode of the intra-level design includes punctuation marks such as periods, commas, dashes,

hyphens etc., and also local marks such as underlined or crossed text (Kostelnick & Roberts, 1998).

Inter-level Design

Intra-level design is made by non-linear components. It helps readers comprehend the text through headings, spatial distribution of the text across the page, and the variety of graphic treatments (e.g. bullets, lines, shadings etc.). Intra-level design makes the text more accessible for the readers. It divides the text into discrete units which are easier for the readers to structure.

The textual mode of the inter-level design includes headings and their size and position, and also numbers within the document. The spatial mode consists of the distribution of the text across the page, and of the division of text into units such as columns, tables etc. Graphic mode of the inter-level design includes bullets in lists, lines between columns, horizontal and vertical lines in tables, and also boxes around text (Kostelnick & Roberts, 1998).

Extra-level Design

Extra-level design consists of various data displays, pictures, icons, and symbols. It includes all the elements that operate outside the main text as autonomous entities with their own visual vocabulary and conventional forms.

Textual mode of the extra-level design includes labels, titles, and legends, as far as concerning data displays. For pictures, it includes all the possible descriptive information, such as labels, call outs, and captions. Spatial mode consists of the key spatial decisions, such as the conventional configuration of data displays (e.g. pie chart, bar chart etc.), selecting sizing and the shape, and the use of perspective. Concerning pictures, it includes the angle of looking. Graphic mode includes shading,

textures, colours of bars, tick marks, gridlines, and also the texture, shading, and details of pictures (Kostelnick & Roberts, 1998).

Supra-level Designs

Supra-level concerns the whole document. It includes the top-down design elements that visually define, structure and unify the entire document. This level often influences the decisions about the previous three levels.

The textual mode of supra-level design includes title pages, chapter and section pages, numbers and tabs signalling breaks in the document, headers, footers, and pagination. The spatial mode consists of arrangement of various elements, page orientation (e.g. horizontal or vertical), page size and its shape, paper thickness, folds, pockets etc. The graphic mode includes all the various marks, icons, colours, linework, and logos that can be found within a document (Kostelnick & Roberts, 1998).

Technical Features of Language Testing

Validity

Validity could be described as the quality which most affects the value of a test. Validity, put simply, is the most important quality of any language test. Every time a test is designed and developed, we have to be certain that it measures only what it is supposed to measure. A test is only valid, if it measures what it is really intended to (Davies et al., 1999; Bachman, 1995; Hughes, 2003). With each language test, there is a question raised: “How much of an individual’s test performance is due to the language abilities we want to measure?” (Bachman, 1995, p. 161).

Validity can be established in a number of different ways. There exist various types of validity and methods of assessing whether a test is valid or not, and it is best

to validate a test in as many ways as possible. Usually, the more important the impact of a test, the more attention we should pay to validity analyse (Alderson, Clapham, & Wall, 1995).

Internal Validity. The first of the two main types of validity, internal validity, relates to studies of the perceived content of the test and its perceived effect. The authors further divide internal validity into three groups (Alderson, Clapham, & Wall, 1995).

Content validity. Content validity is the relevance to and coverage of a certain language domain (Davies et al., 1999). Its main question is whether the content of a test constitutes of a representative sample of the language skills and structures being tested. Content validity is in tight connection with the purpose of the test because the content of a test focused on the same language area would be considerably different for intermediate, upper-intermediate or advanced students. When analysing the content validity of a test, it is essential to have the test specification available (Hughes, 2003).

Test specification is a document which contains the official statement about what the test tests and how it tests it. Test specification creates the basis of elements to be considered for the test, however, not everything in the test specifications will appear in the actual test. This document should contain information about the purpose of the test, usually based on the syllabus of the course or a textbook, information about the students (their age, sex, level of proficiency, native language, cultural background, country of their origin, reason for taking the test etc.), number of sections and papers in the test, text types, language skills tested, language elements tested, number of items in each section, test methods (e.g. multiple choice, gap filling), test rubrics and criteria for assessment. (Alderson, Clapham, & Wall, 1995).

The level of content validity is usually established by comparing the test specifications with the actual test content. This analysis should be carried out by someone who is familiar with language teaching, but who is not directly concerned with the production of the test (Hughes, 2003).

Face validity. Face validity relates to the surface credibility or public acceptability of a test. It is the degree to which a test appears to measure the knowledge and abilities it claims to measure (Hughes, 2003; Davies et al., 1999). Face validity is often misjudged and dismissed as trivial and unscientific because it has to do with appearance rather than with the underlying language construct and is based on intuitive judgement of untrained observers rather than on statistic and scientific analysis. However, if a test does not appear valid, it might not be taken seriously by the test takers and jeopardise the public credibility of a test; that is the main reason face validity has its place in analysing a test as well (Davies et al., 1999; Alderson, Clapham, & Wall, 1995).

Usually, the analysis of test ability includes gathering of data by interviewing the test takers or by asking them to complete a questionnaire about their attitudes and reactions to a test they have just taken or looked at (Alderson, Clapham, & Wall, 1995).

Response validity. The last type of internal validity concerns how individuals respond to test items. Response validity is important because it can often show that although students understand a given passage, they answer incorrectly, and vice versa.

The analysis of response validity is based on gathering introspective data from students. These data can be collected either during the test, which may interfere with the natural response to the test, or after the test, in retrospective. When gathering the

data after the test, it is best to provide the interviewed student with their test or a recording of an oral exam as a support (Alderson, Clapham, & Wall, 1995).

External validity. External validity, the second main type of validity, describes the degree to which results on a test agree with results provided by independent assessment from outside the test. This type of validity is often called criterion validity. External validity is further divided into two groups (Hughes, 2003; Alderson, Clapham, & Wall, 1995).

Concurrent validity. Concurrent validity is established when the comparison of a test scores with some other measure (criterion) for the same candidates taken roughly at the same time. Sometimes the comparison is made on longer and shorter version of the same test (e.g. comparing a 45 min. oral exam to a 10 min. test with a representative sample of language). However, the criterion for concurrent validity is not necessarily a longer test – the test can be also validated against teacher’s assessment of the students (Hughes, 2003).

Predictive validity. Predictive validity is established when the external measures are gathered some time after the actual test has been given. It is a degree to which a teacher can predict students’ future performance. This type of validity is most common with proficiency tests because their purpose is to predict students’ abilities to cope in certain areas (Hughes, 2003; Alderson, Clapham, & Wall, 1995).

The basic rules to increase the validity of a test include explicit specification and use of direct rather than indirect testing wherever feasible. It is also necessary to ensure that the scoring of the test relates directly to what is being tested and that the test is reliable. However, in the case of teacher-made tests, it is unlikely to carry out a full validation and make the test 100% valid (Hughes, 2003).

Reliability

Reliability is another important feature of a language test. This time, it is not concerned with the test as such but rather with its scores. Bachman (1995) defines reliability as the consistency of measures across different times, test forms, raters, and other characteristics. A perfectly reliable score is thus free from errors of measurement.

However, we can never have complete trust in any set of test scores due to many factors which we are incapable to predict; in any testing situation there are several different sources of measurement errors (Hughes, 2003; Bachman, 1995).

Three main factors affect the performance on a language test:

1. Test method facets: testing environment (familiarity, personnel, time, physical conditions), test rubric (test organization, time allocation, instructions), the nature of the input the test taker receives (format, nature of language), the nature of the expected response to the input (format, nature of language, restrictions on response), and the relationship between response and input (reciprocal, nonreciprocal, adaptive).
2. Attributes of the test takers that are not considered part of the language abilities to be measured.
3. Random, largely unpredictable and temporary factors: emotional state of the test taker, changes in the test environment from one day to the next, differences in the way different test administrators carry out their responsibilities.

It should be essential to identify the potential sources of error and try to minimize their effect. By doing so, it is not only possible to minimize the measurement error and increase reliability, but also to satisfy a necessary condition for validity (Bachman, 1995).

The ideal reliability coefficient would be 1, meaning that the test produces precisely the same result. Lado suggests that there are reliability coefficients to be expected for different types of tests: reading – 0.90-0.99, listening – 0.80-0.89, speaking – 0.70-0.79 (as cited in Hughes, 2003, p. 39). In general, the higher the importance of the test, the more focus we should pay to reliability. There are three approaches to estimating reliability:

Internal consistency. Internal consistency concerns with how consistent test takers' performances on the different parts of the test are with each other. To establish the internal consistency of a test, we usually use the split-half method, which means using only one test to get two sets of scores. We divide the test into equal halves and determine the extent to which scores on these two are consistent with each other. It is essential to determine equal halves that are independent of each other (Hughes, 2003; Bachman, 1995). We can either divide the test to first and second half; however, we are not able to apply this method to all tests. Some tests designed as so called 'power tests' usually begin with easier questions and proceed with questions of higher difficulty. Thus, the halves of a test divided in such way would not be equal and the scores would differ. Another possibility is to divide the test into odd and even items, considering the items measure the same ability (Hughes, 2003; Bachman, 1995). We should always try to establish the internal consistency of a test first because if a test is not reliable in its respect, it is unlikely to be reliable to other forms.

Stability. We usually measure stability of a test when the internal consistency of a test does not work (e.g. we are not able to divide the test into equal halves). To measure the stability of a test, we use the test-retest method, meaning that we administer the same test twice and then compute the correlation of the scores. However, this method has several issues. First, the test takers' ability may change

over time due to gaining new pieces of knowledge or the process of ‘unlearning’.

Second, the test takers might remember the test if the second administration follows the first one too soon after, because there is no general period of time between the two administrations. And third, students might be less motivated to write the same test for the second time which also contributes to measurement errors (Hughes, 2003; Bachman, 1995).

Equivalence. When measuring equivalence, we use the alternate forms method. As the name implies, we use two alternatives of the same test (usually A and B) to the same students. The problem with this method is that the alternatives of a test might not be available every time. If we administer the A alternative first and the B alternative second, the students might be influenced by the first form of the test (the practice effect). Thus it is essential to have a counterbalanced design of administration: half of the students gets A form of the test and the other half gets B form as first and vice versa for the second administration (Hughes, 2003; Bachman, 1995).

The approach we chose depends on what we assume the main source of error. With the internal consistency approach, these are the differences in test tasks; stability deals with changes arising as a function of time (e.g. health, state of mind, temperature, audibility, timing etc.), and equivalence focuses on inconsistencies across different forms of tests (Bachman 1995).

To make the test as reliable as possible, we should increase the number of items in the test. The more items in a test, the more reliable the test is. When adding new items to a test, we should have in mind that these items should be independent of each other. The students should not be given a choice, and the range over which possible answers might vary should be restricted. We also should not present students

with items whose meaning is not clear or to which there is an acceptable answer the test administrator did not anticipate. It is understood, that every test should provide clear and explicit instructions. We should use items that permit scoring as objective as possible. We should provide uniform and non-distracting conditions of administration. We should also provide a detailed scoring key and identify students by numbers rather than by their names. Where possible, we should employ multiple, independent scoring. The way the test is laid also contributes to the reliability and so does the legibility of a test (Hughes, 2003).

Usability testing

This section introduces usability testing, which is the main method chosen for the research. The section deals with the definition of usability testing as well as usability in general and its relative terms; further on the method is introduced on a deeper scale, describing the process of conducting a usability test.

Usability

To be able to understand the concept of usability testing, it is important to get familiar with the meaning of usability and the word ‘usable’ as such. When a product is described as usable, users can do what they want to do with the product in the way they expect to be able to use it, without hindrance, hesitation, or questions. In other words, a usable product does not encourage any frustration while the user is using it (Rubin & Chisnell, 2008).

ISO 9241-11 (a standard from the International Organisation for Standardisation covering ergonomics of human-computer interaction) defines usability as “the extent to which a product can be used by specified users to achieve specified goals in a specified context of use with effectiveness, efficiency and satisfaction” (Barum, 2002). *Efficiency* in this case is the quickness with which the

goals set by the user can be accomplished accurately and completely. It is usually a measure of time. *Effectiveness* is the extent to which the product behaves in the way that users expect it to, and the ease with which users can use it to do what they intend. *Satisfaction* is then the user's perceptions, feelings, and opinion of the product. We can usually obtain these through both written and oral questioning (Rubin & Chisnell, 2008). Another important term usability is often connected with is *learnability*, which is actually a part of effectiveness, and has to do with the user's ability to operate the system to some defined level of competence after some predetermined amount and period of training (which often may be no time at all) (Rubin & Chisnell, 2008).

In general, usability is considered a really important feature of any product, because it helps to sell the product, rise the reputation of a company selling the product, as well as lower any type of support and training costs (Dumas & Redish, 1999).

Usability Testing

Knowing what usability means, we can say that the goal of usability testing is to improve the usability of a product (Dumas & Redish, 1999). In a wider viewpoint, it is a term often used to evaluate a product or system by the means of any possible technique. When looking at usability testing from a narrower point of view, it is a process that employs people as testing participants who are representative of the target audience to evaluate the degree to which a product meets specific usability criteria (Rubin & Chisnell, 2008). In other words, participants of usability testing should represent real users performing the same tasks any user in real life would with the given product (Dumas & Redish, 1999).

Nowadays, there exist many different methods that can be used as a part of the usability testing. These can include ethnographic research, focus group research,

walk-throughs, expert or heuristic evaluations, follow-up studies, or varied surveys (Rubin & Chisnell, 2008).

Limitations of usability testing

Even though extremely helpful, usability testing doesn't necessarily ensure that the product will be 100% usable. Testing as such is and always will be an artificial situation, where the very act of conducting a research can affect the results. Those results don't necessarily have to reflect that a product works the way it is supposed to. Last but not least, the participants are rarely really representative of the target group population, because they can be only as representative as our ability to understand and classify the target audience (Rubin & Chisnell, 2008).

The process of usability testing

The process of usability testing is a complex one, ranging from the planning stage, through setting the environment and preparing the documentation, to the evaluation of the results and final report creation.

Planning. When planning for usability testing, we have to think about many areas. Those areas usually include things like establishing an effective team to conduct the test, defining the product issues, setting goals and measurements, establishing the user profile, selecting tasks to test, thinking about how to categorize the results of the test, and writing a test plan (Barnum, 2002).

Test plan. Even though it might seem redundant to create a test plan after taking in consideration all the areas mentioned in the planning section, a test plan is an invaluable document for usability testing. It serves as a blueprint or a guide for the test, helps the testing team to communicate between each other, defines or implies required resources, and provides a focal point for the test. Without a test plan, the

details might get fuzzy and ambiguous; the test plan forces us to approach the test in a systematic manner (Rubin & Chisnell, 2008).

The test plan should include information about the purpose, goals, and objectives of the test; concrete and focused research questions; characteristics of the participants; method of the test; task list; test environment and equipment; the role of the test moderator; the data to be collected and their evaluation measures; as well as the structure of the final report and presentation (Rubin & Chisnell, 2008).

Conducting a test session. When it comes to conducting the test itself, there is a wide range of test variations to choose from. However, the most typical usability test is a one-on-one test conducted with 4-10 participants.

While moderating a test session, the moderator itself can very easily affect what is happening. That is why it is essential to moderate the test impartially, so that the participants cannot sense any preference on the part of the moderator (e.g. through their speech or mannerisms). The moderator should react to the mistakes the same way it reacted to the right answers, and they should never make the participants feel stupid, but rather encourage them. The moderator should not “rescue” the participants when they struggle with something; however, there are situations when the moderator is allowed to assist them – e.g. when the participants feel uncomfortable performing a certain task, when they are exceptionally frustrated and want to give up, or when their action causes a malfunction of the product (Rubin & Chisnell, 2008).

When appropriate, it is advised to use the talk-aloud technique where participants verbally describe what is going on in their head while performing the test tasks, as this technique provides a lot of insights. One of the downfalls of the talk-aloud testing is that the participants still filter their thoughts, and never mention all of their thoughts (Rubin & Chisnell, 2008).

Data analysis. After a test session, the next step is to analyse the data gained. In some situations, it is good to convey a preliminary analysis as soon as possible after the test sessions, trying to pinpoint the worst problems of a given product, so that the designers can start working on their improvement right away. The comprehensive analysis, which usually takes 2-4 weeks after the testing, consist of data compilation (e.g. transferring handwritten notes into a computer, organizing the various types of data), data summary, and data analysis as such (Rubin & Chisnell, 2008).

It is apparent that the technical features of design can be directly affected by the correct or incorrect use of the design rules. By applying the design laws and rules in the right way, it is possible to improve the reliability of a language test as well as its validity. Possible effects of the design of the test can be analysed through use of usability testing which is introduced in a more detailed way in the next chapter.

III. METHODS

This chapter concurs the previous chapter where the method of usability testing was introduced on a theoretical level. In this chapter, the method is described in connection to my research, introducing the tested product, the research questions, as well as depicting the whole process of the usability testing.

Test Plan

Tested product

The research part of the thesis focuses on the examination of the didactic test of the state school-leaving exam (or maturita in Czech) in English, intended for graduating secondary school students of 19 or 20 years of age in autumn 2012. The state school-leaving exam is a relatively new concept in Czech schools; however, the first idea of creating an exam that would unify the exams among high schools in the Czech Republic comes from the late 1990s. The state school-leaving exam first appeared in legislative documents in 2004, with the first tests planned for 2008. However, with the new school law introduced in 2007, the first testing was delayed. Finally, the first phase of state school-leaving exams happened in 2010, and it has been fully implemented since 2012 (MŠMT, 2009).

The state school-leaving exam consists of two parts – the general and school specific parts. The collective part is the same for all secondary schools in the Czech Republic, with its purpose to standardize learning outcomes across schools and to give an opportunity to show a comparison among schools, as well. The general part consists of two mandatory exams. The first exam is in the Czech language and literature, and the second exam is either in a foreign language or mathematics. There is also a maximum of two optional exams in either foreign language or mathematics.

The school specific part, which differs across schools, consists of two or three mandatory exams (the number of these exams is usually decided by the school headmaster with regards to the profile of the field of study) and maximum of two optional exams (Cermat, 2010).

The exam in a foreign language is a so-called comprehensive exam, as it consists of three parts, examining all the key language skills. The purpose of this is not only to test students' overall language knowledge, but also to encourage equal development of all language skills. The language exam consists of a didactic test, written exam, and oral exam. The written part of the test is assigned at all schools in the same way and usually at the same time, while the oral part takes place together with the school specific part of school-leaving exams, in front of an exam committee (Cermat, 2010). The didactic test consists of a listening subtest and a reading subtest, and is written partly in Czech and partly in English, where Czech is used only for instruction and orientation within the document, and English is used for the text of the tasks.

Purpose, goals, objectives

The purpose of the testing was to find out how much the visual design of a document could affect its readers while working with the document. In other words, the goal of the testing was to find out what aesthetic-usability effect the state-leaving exam test in English has on students.

It is a well-known fact that aesthetic designs are perceived as easier to use than less-aesthetic designs. When a design is not as aesthetic as it should, it might result in limited acceptance on the side of users (Lidwell, Holseden, & Butler, 2010).

Research questions

Resulting from the previously mentioned goals of the usability testing, I created the following research questions:

- Does the design of the test support the purpose of the document?
- Are students able to quickly identify sections of the test?
- Are students able to quickly identify individual tasks in the test?
- Are the students able to quickly find the instructions within the test?
- Is the test easy to navigate (e.g. are the individual sections well arranged in regards to their relationship)?
- Is the design of the test consistent throughout the document?

Characteristics of participants

The target group of the state school-leaving exam in English are 19 to 20 year old Czech secondary-school students who would traditionally be the group of participants the usability testing would be conducted with. However, because the purpose of the testing was to analyse the test on the visual level rather than the language one, a group of non-Czech speakers had been chosen for this purpose. This way it was possible to explore the visual chunks of the test without having to deal with the language distractions. It also allowed focusing on the visual design of the test on a more general level.

Participants chosen to be part of the testing come from different countries across the world (e.g. Latvia, Lithuania, Denmark, Estonia, and Thailand). All of them are 19 to 28 year old students.

Testing method

I conducted a sit-by interview kind of usability testing with pre-prepared set of questions. Each of the participants carried out the test individually, only with the test moderator present. The test itself consists of both tasks and questions. During the tasks, the talk-aloud method was used, and participants were encouraged to express their feelings and opinions during the whole testing process freely. Some of the test tasks were timed.

List of test questions and tasks. The test consists of six simple tasks and questions generated with regards to the research questions and testing goals.

1. What do you think this document is?
2. Based on the design of the document, who do you think it is intended for?
Why?
3. How many individual tasks are here in the test?
4. Identify the end of the listening sub-test.
5. Match instructions with their respective tasks.
6. In general, do you find the test easy to navigate? Why / why not?

The first two questions preceded the overall instructions of the test in order to analyse the face validity of the test and the correspondence of the test design with its purpose. Questions 3 to 6 focused on the design of the test in a more detailed way, and explored the ease of use and the ease of navigation throughout the test.

Testing environment

Testing each of the participants individually allowed to let the participants choose the testing location by themselves. It should be a place that makes them feel comfortable, and is not distracting, in order to allow the participants to focus fully on

the test and its tasks. Some of the participants chose to be tested in the library, while others preferred the environment of their own apartments.

Testing equipment

The simple format of the test did not require any special equipment. During the test, the following equipment was used:

- A set of the state school-leaving exam in English working sheets
- Orientation script
- Data collection sheet
- Laptop
- Timer

Once a comprehensive test plan was created, introducing all the necessary elements of the usability testing, it was time to conduct the usability testing with the chosen participants. The next chapter introduces the results gained during the testing, and analyses them further using the design concepts introduced earlier.

IV. RESULTS AND COMMENTARIES

In this section, I introduce and present the results obtained from the usability testing of the state school-leaving exam in English. I am introducing the results in a logical way, according to the questions and their order of appearance in the test. Furthermore, these results are analysed individually, according to the design theories, laws, and rules explained in the previous chapters of this work. At the end, I present an overall analysis of the whole test.

Question #1

The first question of the test (What do you think this document is?) was aimed towards the face validity of the test. It was connected to the research question whether the design of the test supports the purpose of the document, or in other words, whether the document itself is perceived as a language test, because at this point the participants did not know what the document is.

If the design of the document was not appropriate for the purpose of the document (language testing), it could lower its face validity. This could then result in lowering its credibility and public acceptability. It could mean that the students taking the test would not take it seriously enough.

These are the answers I got from the participants of the test:

- 1 It looks like a survey or an exam, because of choosing the correct choice. And exam in English.
- 2 English textbook.
- 3 A test.
- 4 It looks like my English test.
- 5 It's like an exam. These are exam questions or exercise questions, something like that.

- 6 Something about English language. It's about teaching language. Yeah, definitely. It even looks like a preparation for an exam or something. It looks similar to my exam in English.
- 7 Exercise book or probably a test.
- 8 Language test.

These results show that most of the participants (apart from one) identified the document as a test or an exam. Based on the answers, we can see that some of the participants based their answer on their own experience with language testing and identified the document as somewhat similar to their own language tests they had taken. Some of the participants focused mainly on the structure of the document such as the option to choose the right answer and based their answer on this.

Some of the participants hesitated between identifying the document as a test or part of a textbook or exercise book. This might result from the fact that most of the tasks in tests are usually designed similarly to the tasks in a textbook or an exercise book, using the same testing methods. The test itself showcases various testing methods such as ABC questions, true / false exercises, gap filling exercises, matching tasks, and cloze exercises.

We can assume that the face validity of the test is quite strong, as 87.5% of the participants identified it correctly as a test or an exam. This means that the design of the document also corresponds with its purpose.

Question #2

Question #2 was another question aimed at face validity. Through this question, I was trying to find out whether the participants perceive the test appropriate for the age of the students the test was intended to. The aim of this question was to

find out whether the design of the test corresponds with the age of the target group it was designed for.

These are the collected answers:

- 1 17+. The pictures look like it's for kids, but there is more information, so it's probably for adults.
- 2 High school people. There would've been more pictures if it were for kids.
- 3 Kids, or people learning English.
- 4 I would say not kids, because it's very text-heavy. There aren't many pictures. Maybe someone my age and up. Someone who's been to school.
- 5 Students, maybe 10+, because some of the pictures look like for smaller children.
- 6 Probably for kids, because it looks simple, not difficult. 10+.
- 7 High-schoolers.
- 8 I think it can be for kids 10+, but also adults. The pictures look like it's for children, but there are longer texts in the back, so it seems more for adults.

As we can see, most of the participants had trouble deciding whether the target group of the test are children or adults. This was caused mostly by the contrast of the use of simple pictures in the beginning (part 1) of the test, and somewhat text-heavy content towards the end of the document. Five of the participants directly mention the pictures in their answers, saying that they are most likely intended for smaller children (in the age of 10 years old and above, in most cases). However, four out of these five participants deny the test being aimed at children by saying that there is quite a lot of text, or that there would have been more pictures if the test had been intended for children.

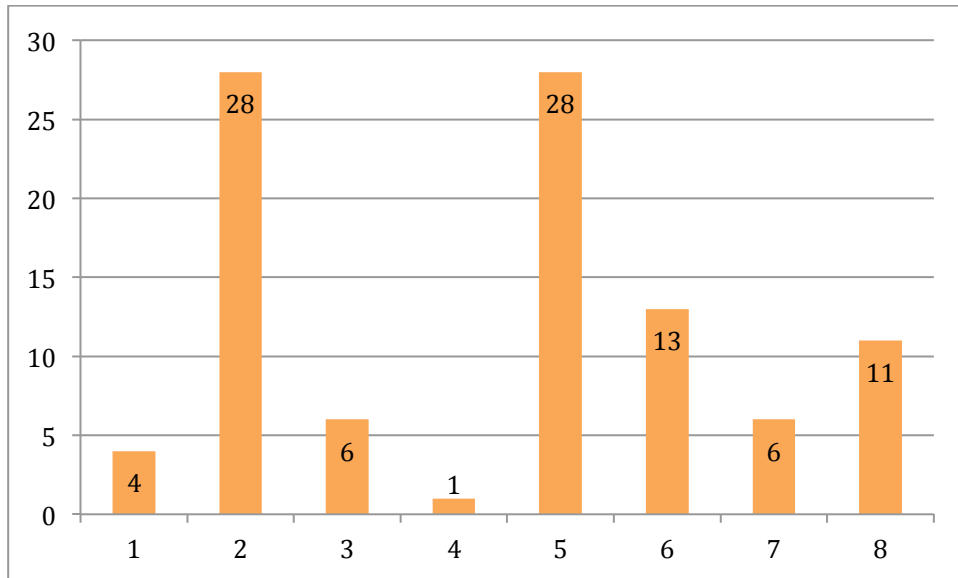
There is a visible clash between choosing children or adults as the target group for the test, which significantly lowers the face validity of the test. The target group of the document is clearly given - 18-19 year old secondary school students; however, this is not intercepted in the overall design of the document. This might result in lowering the credibility of the test. The students taking the test might feel disregarded, which might affect their attitude towards the test, and theoretically even their test score.

This problem could be avoided quite easily. The main reason for the usability test participants to identify the test as targeted towards children was the pictures used in part 1 of the test. This could be avoided by using somewhat more complex, yet still comprehensible pictures, or simply by using photographs instead of illustrations.

Question #3

Question #3 was designed to answer the research questions considering the ease of navigation throughout the document and the ease of identification of individual sections. When a student starts working on a test, it is important to know how many tasks a test consists of in order to be able to plan the time accordingly. That is why the third test question was: How many tasks are there in the test?. This question was timed in order to find out how much time out of the overall 95 minutes the students have to finish the test it takes them to confirm the length of the test.

The average time spent finding out the final number of tasks in the document was 12,125 seconds. The times of individual participants are depicted in Graph 1.



Graph 1. Final times for task #3.

All of the participants answered this question correctly (63 tasks overall). In general, the participants could be divided into three groups based on the way they looked for the information. The first group (50% of participants), with the shortest times to carry out the task, were those looking straight at the end of the test, finding 63 as the number of the very last task on the page. The second group (25% of participants) started going through the document, but eventually stopped, and went to the last page. The third group (25% of participants) browsed the whole document until reaching the last page.

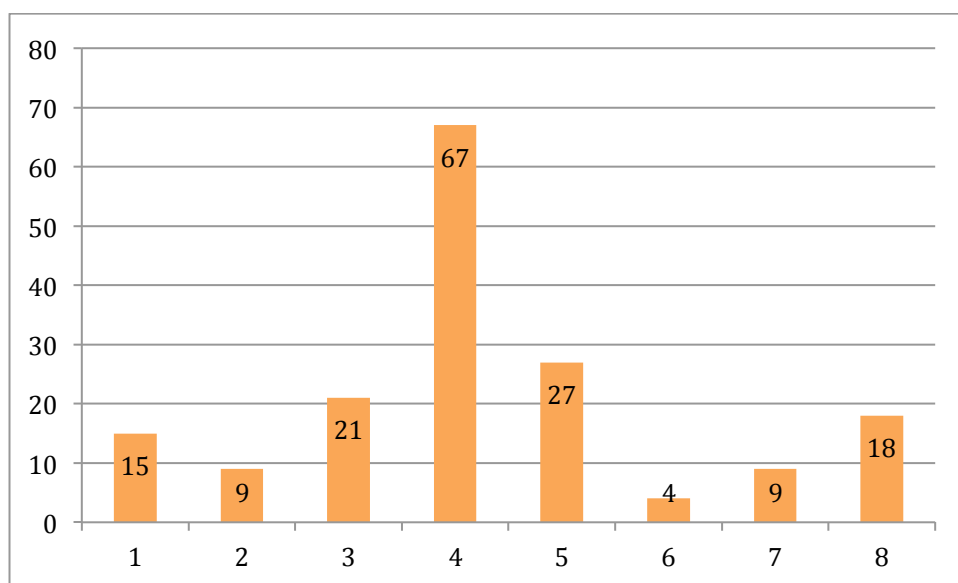
The times to accomplish this task range from 1 second to 28 seconds. No matter which method the participants chose to reach the goal, it takes only about 0.5% of the overall time of the test to finish this task. This shows that the document is quite easy to navigate, thanks to the use of the design rules of repetition and contrast in numbering the exercises throughout the test. All of the exercises are numbered using a bold sans-serif font, which not only helps the test takers spot them easily and distinguish them from the rest of the text, but also makes it easier for the students to orientate within the test.

Question #4

The didactic test of the state school-leaving exam in English consists of a listening part and a reading part. Question number 4 was created to distinguish, whether the division between these two parts is clearly visible and easily identifiable for the test takers. The goal of this task was not only to find out whether the participants are able to identify the division correctly, but also in order to establish whether they are able to do so effectively. That is why this task, same as the previous one, was also timed.

Before the actual task, the participants were introduced to the structure of the test in more depth. They were familiarised with the fact that the test consists of two parts, the first part being a subtest focused on listening and the second part being a subtest focused on reading.

The average time from start to finish (identifying a section of the test the participants considered the division between the two sub-tests) was 21.25 seconds, with the minimum time being 4 seconds, and the maximum time being 67 seconds. The times of individual participants are depicted in Graph 2.



Graph 2. Final times for task #4.

Seven participants out of the eight participants in total identified the subtest division correctly as pages 6 and 7, stating that they chose these two pages, because, apart from the rest of the test, they were blank. This shows the use of contrast within the structure of the test, as the pages of the actual test comprise mostly of text contrary to the dividing pages.

One of the participants, however, identified the division incorrectly as part 6 of the test (starting on page 11). This might have been caused by the different formatting of this particular section (more about section 6 in Question #5).

In general, most of the participants claimed the division section is clearly visible and easily spotted, and thus we can assume that the intended function of the dividing pages works well, and there are no changes needed in this area.

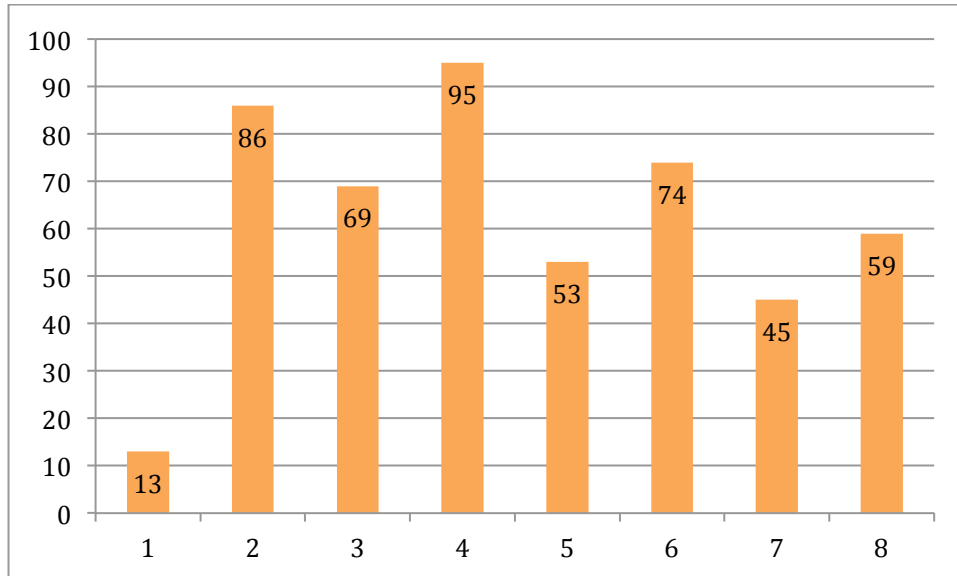
Question #5

Being able to match the instructions in a test with their respective exercises is one of the most important things to be able to do in order to complete a test. That is why the question number five was aimed towards this goal. It is connected with the research questions concerning not only finding the instructions within the test, but also the identification of individual sections of the test, the ease of navigation, and the consistency of design throughout the document.

This task was measuring the number of errors the participants made while matching the instructions and their tasks, and, same as the previous two tasks, it was also timed in order to find out how easy and effectively the test-takers are able to navigate within the test.

The average time it took to go through the whole test and allocate the instructions and the tasks the participants thought were matching is 61.75 seconds. As the test consists of 9 sections of exercises, this counts to about 6.9 seconds per

section. The participants spent about 1% of the time restricted for carrying out the whole didactic test on matching the tasks and their instructions. The times of individual participants are depicted in Graph 3.



Graph 3. Final times for task #5.

Five (62.5%) out of eight participants were able to match the instructions and exercises correctly, while three of them (37.5%) made one or two mistakes during this task. There appeared to be two sections in particular that complicated the process of assigning the right instructions to the right task even for those who managed to carry out task number 5 without errors. These were sections number 6 and number 8. All of the errors made during this task were also concerning only these two sections.

Most of the participants commented that these two sections were confusing for them, and were not sure in what order they should be read. There were multiple reasons for this confusion. One of the reasons is the reverse order of the part with text for reading and the part with options / answers. In all the other exercises, the text usually precedes the options, meaning that the overall consistency of the test and the rule of repetition were broken in sections 6 and 8. In section 6, the continuity is

broken further more by using a different font than in the rest of the test. Most of the test is written with a sans serif font, while a serif font is used solely in section 6.

Furthermore, by positioning the options before the text, the laws of proximity and continuity are broken as well. In both sections, there is a part of the options page blank, which might create an illusion of the end of the exercise. As both of the sections with text occupy the whole page, putting these before the options would create a much stronger sense of continuity.

Several of the test participants expressed their confusion with the order of pages in sections 6 and 8, and some of them were thus forced to use the elements of the textual mode of the supra level of the design as guiding points. These were either the pagination, or the headings. However, one of the participants proposed a change in the headings (particularly in section 8). Their proposal was to either differ the heading of the first page of the section and the headings of the following pages of the same section, or get rid of the headings in the following pages all around. This would create a contrast between the first page and the following pages, making it even easier to navigate the document. This way the test-takers would know which page of the section is the first / main one, without even searching for the instructions.

Question #6

The last question of the usability test was the most subjective question in the test, trying to find out the participants' opinions about the overall ease of navigation of the test. This question is mainly focused on the research question concerning the navigation of the test, but it also deals with the issues of identification of sections and tasks within the test and the consistency of the whole document.

When asked whether they find the test easy to navigate, all of the participants answered yes, giving various reasons for their answer, and mentioning different

means that helped them orientate within the document. In general, the participants said that the text feels like it is built logically, and that the layout somewhat helps them to navigate through the test; however, some of the participants mentioned particular parts of the test that helped them.

One of the participants mentioned that the questions and answers are in bold, which suggests the use of the rules of contrast, repetition, and also the law of similarity. By using a bolder text, these areas are distinguished from the rest of the text, creating a somewhat unified unit within the document.

Other participants said that the options and spaces for answers are easily to be spotted. This, again, shows the use of contrast. The closed answers convey a capital letter in front of them and are considerably shorter in comparison to the regular text. The answers where students are required to write are usually depicted by using a vertical line whose form strongly contrasts with the text.

The position of the instructions was also mentioned as one of the means that helped the participants browse through the document effectively. All of the instructions are placed at the top of the page, using the law of repetition and thus creating a visually balanced design. This fact also strengthens the textual mode of the supra level of the design. Another means of repetition pointed out by the participants was the numbering of the exercises, which is analysed in more depth in Question #3.

The participants also mentioned the spaces between exercises as one of the visual guides. Using the rule of proximity and creating bigger gaps after each exercise makes the test even easier to navigate and it also makes it easier to scan the pages quickly.

Conclusion

The usability testing helped me answer the research questions established during the testing preparation phase. Most of the questions were answered without any problems, and thus we can say that the didactic test of the state school-leaving exam in English is designed quite well. However, two of these questions discovered that particular parts of the test might cause trouble to the test-takers and make it difficult for them to navigate through the exercises.

The very first research question concerning whether the design of the test supports its purpose (language testing) discovered that the test itself is perceived as a language test by most of the participants, and even though it was mistaken for an English textbook, we can say that this shows the test's design corresponds with its function. However, in distinguishing the target group of the test, at first sight the participants had hard time deciding whether it is a test designed for children or adults. As the intended target group of the test are 18-19 year old students, the test itself should look like a test for adolescents or adults in order not to make the students feel underestimated, and in order to keep the face validity, and thus reliability. The factor that made most of the participants think the test was intended for children were the pictures in the first part of the test and their simplicity. This problem could be solved simply by using more complex pictures.

Another problem discovered during the testing concerns the consistency of the test. That is not to say that the test lacks consistency all around. The participants themselves mentioned some re-occurring parts of the test as those helping them with the navigation. These were the instructions being places at the top of the page, and also numbering of sections and exercises. While some of the participants also mentioned headers as one of the visual leads, other participants found them confusing

in sections comprising of two pages, which brings us to the encountered problem. There were two sections in particular which caused problems to most of the participants and made them make mistakes in the task or slowed them down. These were sections 6 and 8. These sections did not follow the rule of consistency and broke it by employing the options before the actual reading part of the exercises. This caused confusion to the participants, because they were unsure about the placement of the individual pages. As mentioned, this problem could be solved by switching the pages of these particular sections. In addition, the headers of the second pages of the two-page exercises should be either changed in order to contrast the header of the first page of these exercises, or deleted in general.

Apart from the problems mentioned above, there were no difficulties experienced with other research questions during the usability testing. The participants were able to identify individual sections, tasks, and instructions within the test. The test was found easy to navigate, both objectively, as the participants had mostly no difficulties carrying out the tasks of the usability test, and subjectively, as the participants themselves found the test comprehensive and easy to navigate. In conclusion, the didactic test of the state school-leaving exam is designed quite well and allowing easy navigation through its parts.

V. IMPLICATIONS

In this chapter, based on the knowledge gained from the usability testing I introduce some general rules which might help teachers with the creation of their own tests in order to make their visual side work for the students rather than against them. I also explain some of the delimitations of my research and its possible weaknesses. Furthermore, I suggest possible ways of extending the research conducted for this work.

Implications for Teaching

The usability testing results showed that some parts of the test's design seems to be of a slightly bigger importance than the rest. Probably the most prominent one of them is the need for consistency throughout the test. It is important to keep the design of the test unified, think about its structure, and not to place any of its elements arbitrarily. The repetition of certain elements is what helps the test-takers orientate within the test, and to know what part of the test they are in at any given moment. One of the elements helping to create a consistently looking test is the use of numbers. The teachers can use numbers in pagination, helping to show the length of the test and also the order of pages. As encountered in the researched test, it is very beneficial to use numbering also for the exercises and even bigger sections of the test (e.g. various sub-tests or parts of the test conveying multiple exercises). The test-takers use these not only as means of navigation throughout the test, but secondary also as means for planning the time during the test.

Another important feature of a test are visible instructions. Instructions as such are one of the most important parts of the test, because without them the students are not able to tell what they are supposed to do. The teachers should not assume that their students know what is expected from them, even though they might have

practiced a certain kind of tasks multiple times before. Thus it is important to always include clear and simple instructions in the test. Considering the design of the test, the instructions should be visible and also clearly distinguished from the individual exercises. The instructions should follow the rule of repetition and contrast. During the testing, it proved useful to place instructions at the top of the page. However, with tests of a smaller scale, where there it is not possible to dedicate an individual page to every single exercise, it is advised to use the same formatting for all the instructions.

The exercises itself should follow the same rules as the instructions. It is important that students are certain as where individual exercises start and where they end. This can be achieved by simple use of the rule of proximity. There should be a visible space before and after each exercise in order to clearly distinguish them from each other. Another possible solution is to apply the rule of contrast and include boxes around the various exercises.

As the test showed, it is important to focus not only on the individual parts and elements within the test, but also on the design of the test as a whole. The teacher should adjust the content of the test not only to the tested skills in order to keep the test valid and reliable, but they should also match the content with its target group. This means that they should use exercises and elements which are appropriate for different ages of students. The usability test showed that using too simple and cartoonish pictures in a test intended for adolescent or even adult students could lower the face validity of the test and make it look like a test for children. This could lower the acceptance of the test in the eyes of its takers. Vice versa, using complex pictures in a test intended for small children could negatively affect both its validity and reliability.

Last but not least, however tempting it might seem to try new things when it comes to language testing, it is better to structure and design the test in a way that students are used to. Designs which look somewhat familiar and which students are able to identify as a design of a test not only adds to the face validity of the test, but also makes it immensely easier for students to work with it.

Limitations of the Research

Even though the usability test proved useful in many areas and helped discover several drawbacks of the didactic test of the state school-leaving exam in English, the research itself had several limitations to it. The foremost one of them, in my opinion, is the fact that there was only one area tested. The research was focused solely on the design of the test and its visual appearance and its possible impacts on the students' performance. However, there are many more areas which can affect the final result of a test. Dedicating all the focus to only one area made the research somewhat lightweight. In order to truly find out whether a test is valid, reliable, and effective, there would have to be a much deeper research carried out. Such kind of research would allow to interconnect the different areas on a much deeper level, and thus bring much clearer results. This fact is also one of the reasons I am unable to establish in general whether the didactic test of the state school-leaving exam in English is designed well or not.

The research conducted for this work cannot be fully generalised, as its implications draw from the results of a usability testing conducted on one particular test. A research study conducted on several different language tests (both commercial and non-commercial) could possibly unfold even more areas of possible improvements.

Another possible drawback of the research is the fact that I am not an expert in usability testing, and thus might not have been able to comprehend the test and its results fully. Usability testing teams usually consist of multiple people, and thus being the only researcher might have been also limiting in a way.

Suggestions for further research

Drawing from the limitations mentioned in the previous section of this chapter, there appears to be one major area suggesting the further development of the research. The research executed in this work was focusing on only one of the aspects of a language test – its visual design. Further research could thus involve some other areas such as the construction of individual tasks and exercises or the way the language test is evaluated.

The research could be extended not only qualitatively, but also quantitatively, including multiple language tests created either by teachers themselves or companies focused on designing tests for schools and other organisations.

A possible next step of such research could be the creation of a general guide book introducing basic rules which might help in creating language tests which are both valid, reliable, effective, and non-constrictive in any way.

VI. CONCLUSION

There are various factors influencing the students' perception of a language test. One of these factors is the visual design. As described in the theoretical part of this thesis, the proper usage of the basic design rules and laws can result in creating a well-structured and overall comprehensible test.

The visual design can not only affect the students' perception of a test, but it can also have an impact on some of the technical features of language testing, such as the validity and reliability of the test, and thus it can affect the students' attitude toward the test, as well as their results.

The goal of this thesis was to analyse the didactic test of the state school-leaving exam in English from autumn 2012 by means of usability testing, in order to reveal possible design issues in the design of the test. The research showed several problems, some of which were causing serious trouble with navigation and orientation within the document.

Based on the findings from usability testing, we could say that consistency and the use of the rule of repetition in the document shows as the most important factor that can affect students' ability to navigate through the document. However, the usage of the other design rules (such as the rule of contrast and the rule of proximity) also help to an indispensable extent.



REFERENCES

- Alderson, J., Clapham, C., & Wall, D. (1995). *Language test construction and evaluation*. Cambridge: Cambridge University Press.
- Ambrose, G., & Harris, P. (2007). *The layout book* (Vol. 2007). Lausanne: AVA Publishing.
- Barnum, C. M. (2001). *Usability testing and research*. Old Tappan: Pearson Education.
- Bachman, L. (1995). *Fundamental considerations in language Testing*. Oxford: Oxford University Press.
- Cermat. (2010). Oficiální stránky nové maturitní zkoušky. Retrieved April 30, 2013, from <http://www.novamaturita.cz/maturita-2013-1404035826.html>
- Davies, A., Brown, A., Elder, C., Hill, K., Lumley, T., & McNamara, T. (1999). *Dictionary of language testing* (p. 284). Cambridge: Cambridge University Press.
- Dumas, J., & Redish, J. (1999). *A practical guide to usability testing*. Portland: Intellect Books.
- Hampe, M., & Konsorski-Lang, S. (2010). *The design of material, organism, and minds: Different understandings of design*. Heidelberg: Springer Verlag.
- Hughes, A. (2003). *Testing for language teachers*. Cambridge: Cambridge University Press.
- Kostelnick, C., & Roberts, D. D. (1998). *Designing visual language: Strategies for professional communicators*. Old Tappan: Pearson Education.
- Landa, R. (2011). *Graphic design solutions*. Boston: Wadsworth.
- Lidwell, L., Holsen, K., & Butler, J. (2010). *Universal principles of design*. Beverly: Rockport Publishers.

- Lupton, E., & Phillips, J. C. (2008). *Graphic design: The new basics*. New York: Princeton Architectural Press.
- MŠMT. (2009). Státní maturita: Nejčastěji pokládané otázky. Retrieved April 30, 2013, from <http://www.msmt.cz/statni-maturita/nejcasteji-pokladane-otazky>
- Rubin, J., & Chisnell, D. (2008). *Handbook of usability testing*. Indianapolis: Wiley Publishing.
- Steinfeld, E., & Maisel, J. (2012). *Universal design: Creating inclusive environments*. Hoboken: John Wiley & Sons.
- Ware, C. (2012). *Information visualization: Perception for design*. Waltham: Elsevier.
- Weinschenk, S. M. (2011). *100 things every designer needs to know about people*. (M. J. Nolan, Ed.). Berkeley: New Riders.
- Williams, R. (2008). *The non-designer's design book*. Berkeley: Peachpit Press.

APPENDICES

Appendix 1: Didactic test for the state school-leaving exam in English (autumn 2012)



ANGLICKÝ JAZYK

základní úroveň obtížnosti AJMZD12C0T04

DIDAKTICKÝ TEST – POSLECH, ČTENÍ A JAZYKOVÁ KOMPETENCE

Maximální bodové hodnocení: 93 bodů
Hranice úspěšnosti: 44 %

1 Základní informace k zadání zkoušky

- Didaktický test obsahuje 63 úloh.
- Časový limit pro řešení didaktického testu je uveden na záznamovém archu.
- Povolené pomůcky: pouze psací potřeby.
- U každé části je uvedena váha části/úlohy v bodech, např.:
5 bodů/1 bod = v celé části můžete získat nejvýše 5 bodů, za jednu správnou odpověď získáte 1 bod.
- U všech úloh je právě jedna odpověď správná.
- Za nesprávnou nebo neuvedenou odpověď se body neodečítají.
- Odpovědi píšete do záznamového archu.
- Poznámky si můžete dělat do testového sešitu, nebudou však předmětem hodnocení.
- Nejednoznačný nebo nečitelný zápis odpovědi bude považován za chybné řešení.

2 Pravidla správného zápisu odpovědi

- Odpovědi zaznamenávejte modrou nebo černou propisovací tužkou, která píše dostatečně silně a nepřerušovaně.
- Hodnoceny budou pouze odpovědi uvedené v záznamovém archu.

2.1 Pokyny k uzavřeným úlohám

- Odpověď, kterou považujete za správnou, zřetelně zakřížkujte v příslušném bílém poli záznamového archu, a to přesně z rohu do rohu dle obrázku.

4

A	B	C	D
<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

- Pokud budete chtít následně zvolit jinou odpověď, zabarvíte pečlivě původně zakřížkované pole a zvolenou odpověď vyznačíte křížkem do nového pole.

4


A	B	C	D
<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>

- Jakýkoli jiný způsob záznamu odpovědi a jejich oprav bude považován za nesprávnou odpověď.
- Pokud zakřížkujete více než jedno pole, bude vaše odpověď považována za nesprávnou.

2.2 Pokyny k otevřeným úlohám

- Odpovědi píšete čitelně do vyznačených bílých polí.

16



- Povoleno je psací i tiskací písmo a číslice.
- Při psaní odpovědi rozlišujte velká a malá písmena.
- Pokud budete chtít následně zvolit jinou odpověď, pak původní odpověď přeškrtněte a novou odpověď запиšte do stejného pole. Vaše odpověď nesmí přesáhnout hranice vyznačeného pole.

Testový sešit neotvírejte, počkejte na pokyn!

© Centrum pro zjišťování výsledků vzdělávání (CERMAT), 2012
Obsah testového sešitu je chráněn autorskými právy. Jakékoli jeho užití, jakož i užití jakékoli jeho části pro komerční účely či pro jejich přímou i nepřímou podporu bez předchozího explicitního písemného souhlasu CERMATU bude ve smyslu obecně závazných právních norem považováno za porušení autorských práv.

1

POSLECH

1. ČÁST

ÚLOHY 1–4

4 body/1 bod

Uslýšíte čtyři krátké nahrávky. Nejprve uslyšíte otázku a poté vyslechnete nahrávku. Na základě vyslechnutých nahrávek vyberte k úlohám 1–4 vždy jeden správný obrázek A–C.

1 What happened to Kyle?

A)



B)



C)



2 What is not on sale?

A)



B)



C)



3 How is the boy going to school?

A)



B)



C)



4 What is the man doing on Saturday?

A)



B)



C)



Uslyšíte rozhovor matky a syna o mobilním telefonu. Na základě vyslechnuté nahrávky rozhodněte, zda jsou tvrzení v úlohách 5–12 pravdivá (P), nebo nepravdivá (N).

- | | P | N |
|--|--------------------------|--------------------------|
| 5 Paul's mobile was damaged by water. | <input type="checkbox"/> | <input type="checkbox"/> |
| 6 Jonathan offered to buy a new mobile for Paul. | <input type="checkbox"/> | <input type="checkbox"/> |
| 7 Paul's mobile was more expensive than Jonathan's. | <input type="checkbox"/> | <input type="checkbox"/> |
| 8 Jonathan's parents are out of the country. | <input type="checkbox"/> | <input type="checkbox"/> |
| 9 Jonathan's grandmother loves watching television. | <input type="checkbox"/> | <input type="checkbox"/> |
| 10 Paul does <u>not</u> remember his friends' phone numbers. | <input type="checkbox"/> | <input type="checkbox"/> |
| 11 Paul is worried about his final exams. | <input type="checkbox"/> | <input type="checkbox"/> |
| 12 Paul's mum thinks that talking to Jonathan's parents will be difficult. | <input type="checkbox"/> | <input type="checkbox"/> |

3. ČÁST**ÚLOHY 13–19****14 bodů/2 body**

Uслыšíte vzkaz, který nechala Lauren svému kamarádovi Johnovi na záznamníku. Na základě vyslechnuté nahrávky **doplňte** informace na vynechaná místa v úlohách 13–19. V odpovědích použijte **nejvýše 3 slova**. Čísla můžete zapisovat číslicemi. První úloha slouží jako vzor (0).

Lauren's friend's name is John (0) Green.

Lauren's brother lives in (13) _____.

Lauren's brother will arrive at (14) _____ o'clock.

Lauren's brother works at a music (15) _____.

Lauren suggests that they meet at the (16) _____ restaurant.

The restaurant is situated next to the library, (17) _____ the park.

The cinema offers cheap tickets for all films on (18) _____.

Lauren's new mobile number is (19) _____.

Uslyšíte čtyři krátké nahrávky. Nejprve uslyšíte otázku a poté vyslechnete nahrávku. Na základě vyslechnutých nahrávek **vyberte** k úlohám 20–23 vždy jednu správnou odpověď **A–C**.

20 What is Karen doing tomorrow afternoon?

- A) She is going to an interview.
- B) She is meeting with her boss.
- C) She is picking up her friend from the airport.

21 Which train will the woman take?

The train that leaves at:

- A) 10:00.
- B) 10:30.
- C) 12:00.

22 Why does the man use public transport?

Because:

- A) it is faster than driving.
- B) it is cheaper than driving.
- C) it is better for the environment.

23 Why is the man calling?

He wants:

- A) to ask Jennifer to babysit.
- B) to have dinner with Jennifer.
- C) to invite Jennifer to a concert.

ZKONTROLUJTE, ZDA JSTE DO ZÁZNAMOVÉHO ARCHU UVEDLI/VŠECHNY ODPOVĚDI.

**NÁSLEDUJE SUBTEST ČTENÍ A JAZYKOVÁ KOMPETENCE.
NEOTÁČEJTE! VYČKEJTE NA POKYN ZADAVATELE!**

© Centrum pro zjišťování výsledků vzdělávání (CERMAT), 2012

7

ČTENÍ A JAZYKOVÁ KOMPETENCE

5. ČÁST

ÚLOHY 24–28

5 bodů / 1 bod

Přečtěte si pět krátkých textů. Na základě informací v textech vyberte k úlohám 24–28 vždy jednu správnou odpověď A–D.

Tom,

Welcome back to work! I hope you're feeling better today. It seems like everyone at the office has been ill recently... I hope I don't catch it! Listen, you don't have to finish your report by 3 o'clock today. The meeting has been cancelled. Not everyone is here today, so the boss decided to meet next week instead. So, mark your calendar! Next Wednesday at 3 p.m. They're supposed to repaint the meeting room this weekend, so I hope it doesn't still smell in there during the meeting!

See you soon,
Janet

(CERMAT)

24 Why is the meeting cancelled?

Because:

- A) Tom has been feeling ill.
- B) some people are absent.
- C) Tom has not finished his report.
- D) the meeting room is being painted.

5th Annual Hip Hop Fest

It's that time of year again, so be sure to get your tickets! Three days of quality hip hop on four different stages. More than 50 performers from around the world, including DJ Price (UK), The VIP Stylers (USA), TicketMachine (Italy), and the incredible DJ Boss (Poland).

When: Friday, Sept. 9 to Sunday, Sept. 11

Where: Bramham Park, Leeds

You can find tickets at all HipMart music stores and at the festival. Both reduced price and VIP tickets this year will only be available online. Visit our e-shop at www.fest11.com. For more information, call 977-9999.

(CERMAT)

25 How can you buy discounted tickets?

- A) by telephone
- B) at the festival
- C) on the website
- D) at music shops

Fast Food Fever

The large fast food breakfast menu is going through a change. Out will be some low sellers such as burgers; in will be a less fattening breakfast menu. Almost every fast food company is moving towards a planned cereal breakfast, which is now being tested. The menu changes could be healthy for customers. People think that fast breakfast foods are unhealthy and consumers want healthy alternatives. The \$1.99 oatmeal breakfast has been welcomed since the start of the test this spring. The test is going very well and customer feedback has been extremely positive. This is the way for fast food restaurants to differentiate themselves from the busy breakfast business right now.

(The Chicago Tribune 26. 8. 2010, upraveno)

26 What does the article inform you about?

- A) a new fast food breakfast menu
- B) an unhealthy fast food breakfast menu
- C) the low cost of fast food breakfast menus
- D) completed testing of a fast food breakfast menu

Gregory Kats, a computer engineering student at a New York university, recently gave passengers on a New York subway train quite a shock. The student was carrying a school project inside his backpack: a model of an elevator, which contained a small battery, wires and a motor. Unfortunately, the model elevator started burning inside his backpack and the fire frightened people around him. Gregory told the travellers that it was only a school assignment, but people thought it was a bomb and ran away in panic. The police soon arrived and questioned him in detail, but they later let him go and Gregory said he was sorry for the trouble.

(www.online-skola.cz, upraveno)

27 Which would be the best title for this article?

- A) Student Finds Bomb on Subway Train
- B) School Project Scares Subway Passengers
- C) Police Rescue Passengers from Elevator Fire
- D) Student Saves People from Burning Subway Train

Grand opening of Sal's BBQ!

From 6 p.m. to midnight, you'll find great food at unbelievable prices. All barbecue sandwiches are half price, steaks and ribs are 25% off, and all non-alcoholic drinks are completely free! For those of you who like a cold beer with your meal, we also offer 5 different kinds of brew on tap¹ at half price, and bottled beers are only 50 cents each. A special kids' corner will be available with games and treats for children, and a rock 'n' roll band will be on site to perform all evening.

Come on down – and bring your friends!

¹ brew on tap: točené pivo

(CERMAT)

28 What will you find at the restaurant opening?

- A) live music
- B) free sandwiches
- C) 25% cheaper soft drinks
- D) bottled beer at half price

Přečtete si tvrzení v úlohách 29–38 a leták s informacemi o přednášce o bambusu. Na základě informací v textu rozhodněte, zda jsou tvrzení pravdivá (P), nebo nepravdivá (N).

	P	N
29 When Butler was a student, he fell through a bamboo floor.	<input type="checkbox"/>	<input type="checkbox"/>
30 Butler will only speak about today's use of bamboo.	<input type="checkbox"/>	<input type="checkbox"/>
31 Bamboo can grow up to 60 centimetres in one day.	<input type="checkbox"/>	<input type="checkbox"/>
32 Bamboo is good for construction because it does <u>not</u> break in bad weather.	<input type="checkbox"/>	<input type="checkbox"/>
33 There will be food at the lecture.	<input type="checkbox"/>	<input type="checkbox"/>
34 Butler will bring some bamboo from his garden in Vietnam.	<input type="checkbox"/>	<input type="checkbox"/>
35 Tickets will <u>not</u> be for sale on July 15.	<input type="checkbox"/>	<input type="checkbox"/>
36 Tickets bought on the website are \$10.	<input type="checkbox"/>	<input type="checkbox"/>
37 Free parking is available opposite the lecture hall.	<input type="checkbox"/>	<input type="checkbox"/>
38 People with campus parking passes can park outside the library.	<input type="checkbox"/>	<input type="checkbox"/>



Bamboo: The World's Most Important Plant

a lecture by Robert Butler

Saturday July 15, 6 to 7 p.m.
Stanley Conference Hall
(Tickets are \$10 each)



About the speaker

Robert Butler is a botanist and agriculturalist who lives in Mississippi, USA, and he's an expert on the bamboo plant. Butler became fascinated by bamboo while studying in Asia. While in Vietnam he lived in a stilt house¹ whose floor was made with very thin strips of bamboo. The first time he saw it, he thought he was going to fall through the floor, but instead the bamboo held his weight quite easily and Butler developed a lifelong interest in the plant.

About the lecture

Most people only know bamboo as the tall plant that pandas eat, but bamboo is much more than just panda food. In this lecture you will learn about the history of bamboo, how it is used today, and how it can be used in the future.

Interesting facts

- There are about 1,500 documented traditional uses for bamboo. It can be used for making anything from fishing baskets and furniture to fences and houses.
- Bamboo can grow to well over 30 metres, and it's one of the fastest growing plants in the world. A Japanese scientist once measured the growth of one variety at 60 centimetres within only 24 hours.
- Bamboo can be found on all continents except Antarctica.
- Because of its strength, the plant is used in many parts of the world for construction. When a skyscraper is built in Hong Kong, for example, bamboo is used as scaffolding² because it is almost as strong as steel, but it is much lighter and easier to transport. Also, the plant is extremely flexible so it can bend in strong winds without breaking – ideal for a country with many severe storms.
- Bamboo shoots³ are also tasty. They are low in fat and calories, as well as a good source of fibre. (A bamboo buffet will be available during the talk, with bamboo taken from Butler's own home garden, so you will be able to try some for yourself!)

Tickets

Tickets will not be sold on the day of the lecture; they must be purchased at least one day before. Tickets are available from either the university ticket office or our website. (Note: If you buy tickets online, you will receive a \$2 discount on each ticket.) For additional information, please call 555-7374 or visit www.agri-uni.edu/lectures.

Note: The free car park opposite the Stanley Conference Hall is currently under construction, but free parking will be provided for everyone behind the nearby Stockman Conference Hall. The car park outside the campus library will also be available – but only for those with campus parking passes. Cars without passes will be removed from this car park.

(www.onestopenglish.com, upraveno)

¹ stilt house: dům na kůlech

² scaffolding: lešení

³ shoots: výhonky

Přečtěte si článek, který pojednává o recenzích restaurací. Na základě informací v textu **vyberte** k úlohám 39–43 vždy jednu správnou odpověď **A–D**.

The Power of Words

Last year, Australians spent about 15 billion Australian dollars eating at restaurants and cafés. Of course, not everyone had great meals because not all of the 40,000 restaurants, cafés and catering services are good. Of the 484,000 people who now work in the hospitality industry, 250,000 are employed in restaurants, and another 55,000 employees will begin working next year. Not all of them will be good either. So how do you separate the good restaurants from the bad? Answer: by reading reviews.

The role of restaurant reviewers in the country has recently been questioned after a restaurant took a newspaper to court. The restaurant said that because the newspaper's review was so bad, it lost all its customers and had to close down. But the newspaper says it's important that reviewers are honest — that's what readers expect. Their purpose is to provide a guide for people who don't want to spend a lot of money on a meal that's not very good. The editor said if they lose the court case, there will be consequences¹ for theatre and book reviewers too. If a reviewer is worried about getting sued² for his opinion, there won't be any reason for having reviews at all.

Interestingly, restaurant owners have said that a negative review isn't always bad. A bad review, they say, can have the power to make a restaurant improve service, which can attract even more customers. They do agree, however, that criticism should always be fair and constructive. A restaurant may need improving, but it doesn't necessarily have to close down.

Do reviewers have too much influence these days? Some have certainly become famous for their humorous comments, which are often negative. It seems to be easier to be funny while saying something bad rather than saying something kind. In addition, the popularity of Internet blogging today means that there are many more hurtful and unfair reviews on the web, while positive blogs are much less common. The anonymous nature of the Internet maybe even encourages rude criticisms.

Of course, we all know that a review can influence some people, but if you're a regular customer of a restaurant, you probably won't stop going because of one bad review. For a restaurant to fail, it must be unpopular with all its customers, not just a single reviewer. It will be interesting to see what the Australian courts have to say about it all.

(www.onestopenglish.com, upraveno)

¹ consequence: následek, důsledek

² sue: zažalovat (u soudu)

- 39 How many people now work at restaurants in Australia?
- A) 40,000
 - B) 55,000
 - C) 250,000
 - D) 494,000
- 40 What does the newspaper say about restaurant reviewers?
- A) They should be careful what they write.
 - B) They help people not to pay for bad meals.
 - C) They are expected to write negative reviews.
 - D) They should be sent to court if they are not honest.
- 41 What do restaurant owners think?
- A) Restaurant reviewers are too powerful.
 - B) Restaurant reviewers are usually unfair.
 - C) A bad review can sometimes be helpful.
 - D) Badly reviewed restaurants should close down.
- 42 What does the text say about reviews nowadays?
- A) The best reviews are anonymous.
 - B) Internet reviews are often negative.
 - C) Bloggers prefer writing nice reviews.
 - D) The number of reviews is decreasing.
- 43 What is the author's opinion in the final paragraph?
- A) Customers decide if a restaurant is successful.
 - B) A restaurant's failure can depend on one review.
 - C) Bad reviews will probably affect regular customers.
 - D) A reviewer has more influence than he or she knows.

Přečtete si informace o pěti lidech, kteří hledají práci, a pracovní nabídky. Na základě informací v textech přiřadíte k úlohám 44–48 vždy jednu pracovní nabídku A–G. Dvě nabídky jsou navíc a nebudou použity.

44 **Laurence** _____

Laurence is a 35-year-old who needs to find a summer job to earn some money. He has a university degree in accounting but he has never used his skills because he hates offices. He is also a vegetarian so he refuses to touch meat. He enjoys outdoor activities, as well as playing with animals and children.

45 **Margo** _____

Margo is a 51-year-old former journalist and housewife who is tired of cooking and taking care of kids. She would like to do something helpful and that makes her feel good. She does not care about money, but she has bad knees so she cannot move quickly or do a lot of physical labour.

46 **Eric** _____

Eric is a 22-year-old artist who is looking for some creative permanent part-time work. He prefers a flexible schedule. His work experience was at a restaurant, which he hated, and a summer job drawing and painting advertising posters for his friend's band.

47 **Rachel** _____

Rachel is a mature, reliable 17-year-old secondary school student who needs to find part-time work to earn money for a summer camping holiday. She is a terrible cook, but she loves animals and often looks after her baby sister when her parents are away.

48 **Elizabeth** _____

Elizabeth is a 25-year-old who graduated from university with degrees in computer technology and web design. Her hobbies include computers, cooking and playing with her cats. She is looking for a well-paid career that makes use of her skills. She has more than 8 years of computer experience.



A)

NOW HIRING

Sam's Steakhouse is looking for a reliable cook to work at our restaurant. Previous restaurant experience isn't necessary, but you should know how to cook and you need to be a team player. The position is part-time, weekends only. However, if you're good enough, you may have the chance to work full-time. We offer our employees excellent pay and free steaks!

B)

PROFESSIONAL WANTED

The Highland Animal Clinic is looking for a person to join our professional team for a full-time, long-term position creating and maintaining our website. Applicants should have a university degree, a minimum of 3 years' computer experience and the ability to work well with other people in a small office. Starting salary is above standard.

C)

POSITION AVAILABLE

Art Weekly is a magazine focused on the world of art. We publish high-quality articles about the latest art, artists, art supplies, etc. We are looking for an accountant with at least 5 years of experience to work full-time in our office doing accounting work for our magazine. Good computer skills are necessary.

D)

MAKE A DIFFERENCE

Do you want to do something meaningful and actually make a difference in someone's life? Now is your chance! Pathfinders Summer Camp is looking for a responsible adult (at least a secondary school graduate) to work with children at our camp this summer. The job is not easy (you will be outside every day and you will be doing a lot of physical activities like hiking), but it is rewarding with good pay. It's perfect for anyone who wants to escape computers, telephones, and the stress of city life!

E)

LOOKING FOR EXPERIENCE?

The Williamsburg Post is a small-town newspaper in need of a few part-time artists to help us add content to our website. Your job will be to draw cartoons and pictures to go with our articles. This can even be done from home so you can choose when to work. Applicants should be imaginative and able to create interesting pictures. Some previous experience is necessary.

F)

SEEKING RESPONSIBLE HELP

Toddler Daycare is looking for a responsible young person to work at our childcare centre at weekends. Applicants with childcare or babysitting experience are preferred but we will consider anyone who loves working with young children and dogs because we have a small dog here that the children can play with too. Important: Our kids are very active, so you must be able to run around with them!

G)

WE NEED YOU!

Critter Comfort is a non-profit organisation focused on taking care of injured or lost animals. We are looking for responsible volunteers of any age to help out with a variety of jobs – anything from answering phones to feeding dogs or bathing cats. With us it's never boring! So even though you won't get paid, you'll find the work is very rewarding. If you'd like to volunteer, please call us anytime between 10 a.m. and 5 p.m.

(CERMAT)

Přečtěte si článek o festivalu v Edinburghu. Na základě textu **vyberte** k úlohám 49–63 vždy jednu správnou odpověď **A–C**.

The Edinburgh Festival

Edinburgh is a popular place to visit any time of the year but the best time to come is the month of August when the city traditionally (49) _____ into a gigantic open-air arts festival. The Edinburgh Festival Fringe – simply (50) _____ as “The Fringe” – is the biggest open-air arts festival in the world. The Fringe flows into the streets of Edinburgh so that everywhere you (51) _____, there’s a performance going on.

However, the popularity of the Fringe Festival (52) _____ make a trip to Edinburgh in August quite complicated – not to mention expensive. The hotels and hostels book up very quickly and even if you decide (53) _____ a room in advance, you will find out that the room prices are (54) _____ than usual. But perhaps the biggest cost of your trip to Edinburgh – especially if you are travelling from far away – is going to be your plane (55) _____. That’s because The Fringe is a really big attraction for thousands of people (56) _____ takes place in the middle of the high tourist (57) _____ in Europe. That’s two counts against anyone who’s a budget traveller¹.

Still, for those of you who are interested (58) _____ visiting the Edinburgh Festival Fringe this year, there are many (59) _____ to get a discount on your travel. Even the online travel magazine Itchy Feet last month (60) _____ their readers about some of them in a series of articles about Edinburgh. On the website, you will also find practical (61) _____ about public transport, comfortable accommodation and good restaurants in (62) _____ area.

With more than 2,500 events on the programme, this year’s Edinburgh Festival Fringe is going to be the largest in (63) _____ history so if you love the arts, then don’t hesitate to come to Edinburgh in August.

(www.cheapticketlinks.org, upraveno)

¹ budget traveller: cestovatel s omezeným rozpočtem

- | | | | |
|----|-------------|-----------------|-----------------|
| 49 | A) changes | B) changed | C) will change |
| 50 | A) know | B) known | C) knows |
| 51 | A) follow | B) look | C) watch |
| 52 | A) should | B) can | C) has to |
| 53 | A) reserve | B) to reserve | C) reserving |
| 54 | A) higher | B) high | C) the highest |
| 55 | A) ticket | B) bill | C) receipt |
| 56 | A) when | B) who | C) which |
| 57 | A) season | B) term | C) time |
| 58 | A) with | B) about | C) in |
| 59 | A) journeys | B) trips | C) ways |
| 60 | A) informed | B) has informed | C) had informed |
| 61 | A) show | B) promise | C) advice |
| 62 | A) the | B) — | C) an |
| 63 | A) its | B) her | C) his |

ZKONTROLUJTE, ZDA JSTE DO ZÁZNAMOVÉHO ARCHU UVEDLI/A VŠECHNY ODPOVĚDI.

Appendix 2: Data collection chart

Question	Time	Answer	
What do you think this document is?			
Based on the design of the document, who do you think it is intended for? Why?			
How many individual tasks are there in the test?		Answer	Commentaries
Identify the end of the listening sub-test.		Answer	Commentaries
		<input type="checkbox"/> correct <input type="checkbox"/> incorrect	
Match instructions with their respective tasks.		Errors	Commentaries
In general, do you find the test easy to navigate? Why / why not?			

Appendix 3: Data collection chart from participant #1

Question	Time	Answer	
What do you think this document is?		It looks like a survey or an exam, because of choosing the correct choice. And exam in English.	
Based on he design of the document, who do you think it is intended for? Why?		17+. The pictures look like it's for kids, but there is more information, so it's probably for adults.	
How many individual tasks are there in the test?	00:04	Answer	Commentaries
		63	* went straight to the end
Identify the end of the listening sub-test.	00:15	Answer	Commentaries
		[x] correct [] incorrect	Page 6 or 7, because they are blank.
Match instructions with their respective tasks.	00:13	Errors	Commentaries
		2	* errors in sections #6 and #8
In general, do you find the test easy to navigate? Why / why not?		Yes. Questions and answers are in bold, the ABC options and spaces for answers are easily spotted. Just #6 felt out of place.	

Appendix 4: Data collection chart from participant #2

Question	Time	Answer	
What do you think this document is?		English textbook.	
Based on the design of the document, who do you think it is intended for? Why?		High school people. There would've been more pictures if it was for kids.	
How many individual tasks are there in the test?	00:28	Answer	Commentaries
		63	* went through the test
Identify the end of the listening sub-test.	00:09	Answer	Commentaries
		<input checked="" type="checkbox"/> correct <input type="checkbox"/> incorrect	
Match instructions with their respective tasks.	01:26	Errors	Commentaries
		0	* hesitation with section 6
In general, do you find the test easy to navigate? Why / why not?		Yes. Visually, I can tell where the sections start, mostly by the headings.	

Appendix 5: Data collection chart from participant #3

Question	Time	Answer	
What do you think this document is?		A test.	
Based on the design of the document, who do you think it is intended for? Why?		Kids, or people learning English.	
How many individual tasks are there in the test?	00:06	Answer	Commentaries
		63	* went straight to the end of the test
Identify the end of the listening sub-test.	00:21	Answer	Commentaries
		<input type="checkbox"/> correct <input checked="" type="checkbox"/> incorrect	* section 6 identified as the start of the reading sub-test
Match instructions with their respective tasks.	01:09	Errors	Commentaries
		2	* errors in section #6 and #8
In general, do you find the test easy to navigate? Why / why not?		Yes. It is easy, except section #6 and #8. It looks like a test. The header on second page of section #8 is confusing.	

Appendix 6: Data collection chart from participant #4

Question	Time	Answer	
What do you think this document is?		It kinda looks like my English test.	
Based on the design of the document, who do you think it is intended for? Why?		I would say not kids, because it's very text-heavy. There aren't many pictures. Maybe someone my age and up (20+). Someone who's been to school.	
How many individual tasks are there in the test?	00:01	Answer	Commentaries
		63	* knew the answer by heart thanks to the previous browsing
Identify the end of the listening sub-test.	01:07	Answer	Commentaries
		[x] correct [] incorrect	
Match instructions with their respective tasks.	01:35	Errors	Commentaries
		0	* hesitation with part 6 * hesitation with part 8 I suppose these go together, but I don't know in what order. If I didn't have the page numbers, I wouldn't know.
In general, do you find the test easy to navigate? Why / why not?		Yes, except for that one part (shows section #8). Because it's built logically, especially part #5.	

Appendix 7: Data collection chart from participant #5

Question	Time	Answer	
What do you think this document is?		It's like an exam. These are exam questions or exercise questions, something like that.	
Based on the design of the document, who do you think it is intended for? Why?		Students, maybe 10+, because some of the pictures look like for smaller children.	
How many individual tasks are there in the test?	00:28	Answer	Commentaries
		63	* went through the test
Identify the end of the listening sub-test.	00:27	Answer	Commentaries
		[x] correct [] incorrect	
Match instructions with their respective tasks.	00:53	Errors	Commentaries
		0	
In general, do you find the test easy to navigate? Why / why not?		Yes. Because the layout helps you to know what you are supposed to do.	

Appendix 8: Data collection chart from participant #6

Question	Time	Answer	
What do you think this document is?		Something about English language. It's about teaching language. Yeah, definitely. It even looks like a preparation for an exam or something. It looks similar to my exam in English.	
Based on the design of the document, who do you think it is intended for? Why?		Probably for kids, because it looks simple, not difficult. 10+.	
How many individual tasks are there in the test?	00:13	Answer	Commentaries
		63	* went through the test and after a while straight to the end
Identify the end of the listening sub-test.	00:04	Answer	Commentaries
		[x] correct [] incorrect	Page 7.
Match instructions with their respective tasks.	01:14	Errors	Commentaries
		1	* error in section #6
In general, do you find the test easy to navigate? Why / why not?		Yes. The instructions are on the top of the page, the new tasks usually start on a new page, it is obvious.	

Appendix 9: Data collection chart from participant #7

Question	Time	Answer	
What do you think this document is?		Exercise book or probably a test.	
Based on the design of the document, who do you think it is intended for? Why?		High-schoolers.	
How many individual tasks are there in the test?	00:06	Answer	Commentaries
		63	* went straight to the end
Identify the end of the listening sub-test.	00:09	Answer	Commentaries
		<input checked="" type="checkbox"/> correct <input type="checkbox"/> incorrect	It's on the blank pages.
Match instructions with their respective tasks.	00:45	Errors	Commentaries
		0	* hesitation with section #6
In general, do you find the test easy to navigate? Why / why not?		Yes. The exercises are usually on a page or two, and they are numbered.	

Appendix 10: Data collection chart from participant #8

Question	Time	Answer	
What do you think this document is?		Language test.	
Based on the design of the document, who do you think it is intended for? Why?		I think it can be for kids 10+, but also adults. The pictures look like it's for children, but there are longer texts in the back, so it seems more for adults.	
How many individual tasks are there in the test?	00:11	Answer	Commentaries
		63	* went through and after a while straight to the end
Identify the end of the listening sub-test.	00:18	Answer	Commentaries
		[x] correct [] incorrect	
Match instructions with their respective tasks.	00:59	Errors	Commentaries
		0	* problems with section #6 I'm not sure where this belongs. But probably like this (correctly), because it would be weird to have two long texts together (talking about texts in sections #6 and #7).
In general, do you find the test easy to navigate? Why / why not?		Yes, except for that one exercise (#6). There are spaces between the exercises, and they are numbered. The spaces for answers are visible.	

SUMMARY IN CZECH

Tato diplomová práce se zabývá možným dopadem vizuálního designu jazykového testu na vnímání tohoto testu studenty. Práce poskytuje informace o základních pravidlech a zákonech designu, a dále rozebírá jejich využití v didaktickém testu anglického jazyka vytvořeného pro státní maturity. Cílem výzkumu prováděného pomocí testování použitelnosti bylo objevit chyby v designu a navrhnout jejich případná řešení.