



Online Language Assessment

Bachelor Thesis

Bachelor of Science

as part of the Study

Software&Information Engineering

Nikola Gospavic

01427966

Institute for Information Systems Engineering

Research Unit E-Commerce

Faculty of Informatics Technical University Vienna

Mentor: Jürgen Dorn

Wien, 17.01.2020

Declaration on the constitution of the work

Nikola Gospavic
1040 Wien

I hereby declare that I have written this thesis independently, that I have stated the sources and resources used in full, and that I have all of the materials used in this thesis, including tables, maps and illustrations, whether they are from the Internet or other works, properly and clearly cited.

(Place, Date)

(Signature of the author)

Abstract

The following thesis represents a scientific assessment and the development process description of an online tool for language knowledge testing. This thesis culminated with the development of a usable online tool for assessing an individual's language skills, that can be used in a university environment. The main usage of the module will be in situations where a large group of students needs to be evaluated for their knowledge of a certain language, quickly and as accurately as possible. The motivation for this thesis has come from situations in university lectures, where it needs to be decided what language to use for a given lecture, given the ever-growing foreign student presence at all major universities, including our own.

There are many factors to take into consideration, such as grammar, vocabulary and general comprehension of a certain language. All of this needs to fit into a reasonable time frame of expected completion of a single test-run. The module will be developed with the possibility to expand it with any given language. The main language that is to be implemented for assessment is English. After a group is done with the test, the lecturers should be able to know, from the results, if they are able, and to what extent, to follow up in a university-level lecture in the given language.

An extensive exploration of existing thesis needs to be done, in the context of language knowledge assessment and existing online language tests. After that, a language test will be designed, for the English language. The design of the assessment test, will also be tested for consistency in results and accuracy. Finally, the test will be incorporated into a usable online tool, with the development process explained in detail.

Table of contents

Table of Contents.....	1
Figures.....	2
1 Introduction.....	3
1.1 Scope of the Problem.....	8
1.2 Motivation and Goal.....	11
2 State of the Art in Online Language Assessment.....	13
2.1 Definition of the Problem.....	13
2.2 Existing Language Assessment Frameworks.....	17
2.3 Dialang - Language Assessment Framework.....	24
2.4 Research Summary.....	29
3 Solution.....	30
3.1 Technology used.....	30
3.2 Solution.....	34
3.3 Question Design.....	37
4 Conclusion.....	40
5 References.....	41

Figures

1 Introduction

Throughout history, universities have been the turning points in the lives of the most influential humans ever to be, where different ideas are exchanged, where cultures meet, where minds are shaped. In the modern world universities have become increasingly multicultural, and one of the biggest meeting points for a vast and diverse group of people from different cultures and different backgrounds.

In the past half century, more than ever, there has been an influx of more international students to all higher education institutions. This is largely due to more and more countries participating in the global free economy, but also because many higher education institutions are actively increasing their recruitment efforts of international students. Aside from diverse cultural and diplomatic contributions, the hosting countries also usually experience an economic contribution. Domestic students also engage more in challenging their own ideas and beliefs, when exposed to perspectives that are sometimes drastically different than their own, such as with international students. This has shown to be of great importance in the overall intellectual development of an individual. Therefore, in the international flow of scientific talent, nations are investing more, because they are getting more dependent on the migrations of professionals and students, that will later boost the economic growth. Even if the international students choose to leave the hosting country after the studies, the hosting country can still benefit in many ways. This is manifested in a way that students are usually left with a well-being sense toward the hosting country, which can lead to later friendly economic and diplomatic engagements. This model of conduct towards talent has proven as successful over the years and many countries are starting to stimulate foreign students to study at their universities. International students have been in an area of neglect for a long time, but now started getting vast attention. Failing to recognize and plan ahead without the international student population can lead to program failures in academic disciplines, and have bigger consequences for a nation [1].

One big factor to mention here, is the problem with the domestic supply of students. In the United States of America, and in other similarly developed nations, the national rate of citizens who finish high school is diminishing. The older generations of nationals have overall had better school performance and filled a larger percentage of the university enrollments, especially in the STEM fields. Nowadays, the highly competitive environment of the STEM fields, especially the highly selective programs, has been largely filled with international students, to fill the voids [2].

We can thus deduct that a driving factor of a developed nations educated youth, especially in the STEM fields, are international students, who tend to stay in the hosting country after the studies.

Of course, the main language used for communication in higher education institutions, is English. In most undergraduate and post-graduate education in the world, and even doctoral research, English is used. Because of this, more and more countries and universities are offering graduate programs that are entirely taught in English. In Europe, the percentage of international students that are enrolled in universities varies from 10-30% , while in the case of research and doctoral studies, the percentages go up, and range somewhere between 30-40%. The presence of foreign students is especially felt in STEM fields at the undergraduate level. To enroll to an undergraduate study, public as well as private educational institutions have relaxed their requirements for foreign students, followed by the hosting countries usually relaxing the requirements for obtaining visa's or other means of legalities. The effort to attract foreign students is even more felt in competitor countries, which are creating more relevant curricula and programs for foreign students. For example, institutions like the "Ecole des Mines" (Paris Tech) are offering English language master's programs that are specifically targeted to attract international attention. Other famous educational institutions are following the example, by leading the offer in the English-language taught MBA degree market [2].

Becoming aware of the roles and responsibilities of different levels of higher education, is one of the first steps into creating a successful learning environment. This is especially important for international students, as they need to participate in activities just like national students. Fee and Radford suggest that there is a downward shift in domestic student numbers, increased skilled immigration, followed by an increased demand globally for higher education. Student satisfaction is one of the main goals of an university [3]. The creation of great value education has many different factors involved. A study at University of the Incarnate Word in San Antonio, Texas, says that

international student experience challenges with integration, services, finances, social activities, health care, orientation, housing, and various dealings with bureaucracy and the authorities [4].

Let's consider one of the main problems of international students and their most difficult adjustment areas:

Faculty Academically Detrimental Behaviors	Students Perceptions of Most Difficult Adjustment Areas
1. Not participating in class	1. Social isolation
2. Not asking for clarification	2. Language skills
3. Sitting only with international students	3. Knowing norms, rules, and regulations
4. Studying only with international students	4. Overcoming stereotypes
5. Breach of ethical standards of scholarship	5. Transportation
	6. Weather differences
	7. Food differences
	8. Oral presentation assignments
	9. Clothing norms
	10. Personal finances

Table 1. *Most difficult adjustment areas for foreign students [4]*

As can be seen, among other difficult adjustment areas, language skills is one of the perceived by students. Language barriers are one of the most often causes of difficulty in making domestic friends, having feelings of discrimination by domestic students and similar problems. One of the causes of language barriers deteriorating slower, is the fact that international students tend to be friends with other international students. A language difficulty is often a signal for English comprehension issues, because English is the most used medium for communication of all. An English language difficulty may cause poor communication not only with the professors but with other students, including domestic students, as well. This affects the entire education experience for foreign students. It can clearly be seen why language competency, mostly in English, is a critical part of a foreign students success, even before admission, and certainly during the studies. It is worth to mention, that besides

English, other languages are also being widely used, such as German, French, Spanish and Chinese [4].

Taking all of this into account, a logical conclusion is, that the use of foreign languages, especially English, has spiked in the last several decades. Ever increasing foreign-student presence has prompted many universities to adapt their curriculum to a more English-friendly or entirely English held basis. One of the issues in teaching such a diverse group of students, is not knowing their level of comprehension of the target language. The teaching staff, has to, at least to some degree, determine in which language and on what level the lectures can be held. The most straight forward way of doing this, is assessing the language comprehension of students before the first lectures have begun. After that, when the teaching staff has determined the foreign language proficiency of their students, the lecture, and future lectures, can be planned in accordance to the language knowledge levels of the students.

Assessing someone's level of knowledge of a certain language can be, in many ways, tricky. There are many factors to take into consideration, such as grammar, vocabulary and general comprehension of a certain language. It only becomes more difficult when larger groups of students need to be evaluated. When students arrive for the first time at a school, their language proficiency needs to be assessed to identify their needs and inform the instructor so they can plan ahead. An usual approach to this is either asking the students and counting hands, or giving out paper sheet test for the students to fill out. However these approaches take time and are prone to errors. There are many issues in correctly, or at least to some degree, evaluate one's language comprehension and proficiency. Especially when it is required to do so in a short period of time and with relative simplicity. In this thesis, the English knowledge of a person will be evaluated in a university setting. The evaluation is therefore, usually, of a group of students that undergo a certain lecture. Different key factors are present, which are influencing the assessment of one's language knowledge, and need to be taken into account, such as varying levels of proficiency in both the target language, as well as the native language . Adding to this are also one's educational background and the underlying degrees of exposure to things such as standardized testing. It could be the case that some students may have never seen a multiple-choice or constructed-response question [4].

In this thesis, a solution for a quick and easy language assessment of a classroom of students will be developed. The solution will be developed as a simple online tool with the help of existing technologies.

The next part of this thesis will analyze the different ways to approach the language assessment problem, and how to develop a solution that can be used in an university setting. I will try to analyze the problem as best I can, through research and also interviews with foreign students. The problem solution will be developed in accordance to the industry standards of the chosen solution.

1.1 Scope of the Problem

In order to understand the problem of assessing the language knowledge of a student in a university setting, a research in existing work needs to be done, and the best possible solution to incorporate into the online tool needs to be done as well. Further research into the technologies needed for the development of the online tool will be done in this thesis, as well as research into existing online technologies that serve the same purpose.

As already mentioned, there are many factors to consider when designing a test. In the case of this thesis, it needs to be as short as possible and as precise as possible. Also, it needs to minimize the test-takers anxiety, which will be achieved by making it as simple as possible.

One of the most important factor that needs to be taken into account, in regards to designing a test, is the psychological factor. Anxiety is one such factor. Anxiety is an uncontrolled state of emotion that brings nervousness about something uncertain to us. It has been shown that trait anxiety, that we can be genetically predisposed to have, can be exhibited in many different circumstances. One other type of anxiety is situational anxiety, which can be exhibited only in certain specific situations. One of the best-known forms of state anxiety, is test anxiety. Another famous form of anxiety is the so called “foreign language anxiety”. This sort of anxiety is very often exhibited by people who are in the process of learning a foreign language. Horwitz and Cope identified three major components of foreign language anxiety [5].

1. Communication apprehension (inability to express mature thoughts and ideas)
2. Fear of negative evaluation (a need to make a positive social impression)
3. Test anxiety (fear of academic evaluation)

Considering these three components in the scope of development of anxiety in students, an effort has to be made to try and minimize their effects in this online assessment tool. In this case, a student’s foreign language knowledge will (usually) be tested. This will lead to even more possibilities to develop anxiety, as foreign language and test anxiety are two separate types of anxiety. Teachers can usually make a

difference and reduce the learner's anxiety, in the form of enhancing their confidence and encouraging their overall involvement in the classroom. This has been proven to reduce the effects of the mentioned forms of anxiety [6].

Student involvement in the classroom plays a big part in creating a comfortable atmosphere. Matsuda and Gobel say that using activities such as paired work, group work, games and roles, changes the atmosphere significantly, and makes the students feel more relaxed [7]. This leads to lower levels of anxiety. When students are involved in the explanation of the use and evaluation of specific educational strategies, their performance improves. Phillips suggests, that teachers should directly discuss the language issues of the students, because this will reassure them that those feelings are normal [8]. Language learning is a huge process that takes a long time, and a realistic picture should be painted. In the end, the study of Salehi and Marefat showed, that both language and test anxiety negatively relate to foreign language test performance with students [9].

A big effect on the forming anxiety during tests, is the time factor. Many people feel pressured to show their knowledge when there is a time constraint present. In addition to this, learners can get a fear of the grading process. The predominant fear in regards to testing, is the fear of bad grading and the negative consequences of receiving bad grading. A research conducted in 2006 by Dr. Selami in Istanbul, on 114 students in the English Language Department of Necatibey Education Faculty of Balikesir University in Turkey, was made in order to determine the level of test anxiety, what provoked it, what the effects are and what can be done to prevent it. The research was conducted during an English language test, and it showed that almost all test takers usually have test anxiety. The students also reported forgetting what they know when under pressure by time limits and the feel of trouble before the test. One interesting finding here was that many students showed anxiety because the thought that the other students were doing better than them on the test. There were many factors mentioned, that prevent students from showing their real knowledge on the test. One of the most interesting findings was, that besides the usual reasons for test anxiety, participants reported that they did not like being compared to others, in a test environment. For these students, the main reason for test anxiety, was lack of content validity [10].

In the test that was conducted by Dr. Selami in 2006, it was shown that female students feel less confident and relaxed than male students, even when they are prepared well for the test. One interesting thing here was that the female students were predominantly prepared better for the test, than male students. But, all in all, the higher

the scores of the students was, the less the level of anxiety was shown, meaning that those who scored more, felt more confident and relaxed, than those who scored lower. Students who received lesser grades, reported thinking about things unrelated to the test, during the test, much more than students who were prepared better. The ones who experienced the most test anxiety, reported physical problems, such as headaches, stomach pain and increased heart rates during and before the test [10].

As in regards to the way of lowering test anxiety, there were many different answers in Dr. Selami's research. The students made remarks on that, how the teachers can greatly reduce stress and test anxiety by introducing them to the test, handing out trial versions of the actual test and informing them of the goals of the examination. Negative comments and remarks during and before the test should also be avoided. All of this can help show the students that the test is not there to show authority, sanction or discipline [10].

This prompts the question, how much effort should actually be put into reducing test anxiety and how important is this factor in the overall educational sector? Cold test environments can, with relative ease, be made into warmer ones, with words and encouragement. In many cases, test anxiety can simply be avoided by using ongoing evaluation methods with assignments or projects to evaluate, language proficiency in our case. Teachers should be aware of students' anxiety and that it plays a huge role in overall education, and know, at least to a degree, how to prevent it.

Important factors that need to be incorporated into designing an online language assessment tool are test validity, time limits, test format and length, the environment and the instructions that will be given to students. Negative evaluation and comments, should be avoided, to prevent bad experiences that could cause test anxiety. This is especially important in the case of this thesis, because the use case of the online tool will usually involve students new to the course or maybe even new to the whole university setting. Objective scoring methods will be used. Teachers should inform the students of the test contents, the aim and test format, to create a viable and "warmer" atmosphere.

The next part in this chapter will define the motivation and exact goal of this thesis.

1.2 Motivation and goal

The motivation behind this thesis comes mainly from the requirement of an online language knowledge assessment tool, at the Technical University of Vienna. As many international students join various classes at the TU Vienna, it is somewhat hard to grasp, what kind of languages they all speak, and if the lecture should therefore be held in German or English. Even after deciding which of these two suits the student group, it is important to know to which degree the students are knowledgeable of a certain language, so that the appointed professor can plan and write the lectures accordingly. The idea is that this simple tool could also be used at other universities, where a similar use-case appears. Young generations of students are mostly very familiar with English, and, most professors do speak at least one secondary language in addition to their primary language, as well. All of this needs to be taken into account, when designing a language knowledge test, and also its incorporation into an online tool.

The first main usage will be for assessing the language knowledge of a class of students before a lecture. The second usage will be in courses where teams of five students will be required to work together. In this team group setting, the tool will be used by a professor, to evaluate how important language knowledge and skills, are important, in relevance for success in these kinds of projects. This question of language knowledge relevance to a students success in the studies, has been raised before, and is a theme for dispute at many different universities, and in many different fields.

It is important to state, that an even larger pool of international students is expected in the foreseeable future. Of course, this does not mean that the TU Vienna will ever stop using German as the main language of it's courses, as that is the official language of its country, Austria. But, judging by the current situation, most exchange students speak little or no German. So it can pretty much be said, that the use-case of such a tool, or similar tools, is only expected to broaden.

An efficient and elegant solution however, which can give results in the appointed time frame, is not expected to be 100% accurate. The accuracy will however

be enough of an estimate, to conclude with assurance that the students will be able to keep the pace of chosen lecture if it fits with their test results for a certain language. Hopefully this thesis and the research done will aid in some other future tools that can possibly be much more complex. The culmination is of course, a usable tool that will help the teaching staff, to some extent, in dealing with students who do not speak the local language, at the Technical University of Vienna.

2 State of the Art in Online Language Assessment

2.1 Definition of the Problem

The problems that are faced when assessing someone's knowledge of a certain language have been mentioned in the previous chapters. The main things that need to be assessed, are grammar, vocabulary and general comprehension. To fit all of this into a short test time frame, the use and/or modification of existing tools available for language knowledge assessment needs to be done. How is practicality achieved? Firstly, the main objectives are determined, the teacher's and students' time constraints. After that, a clear path of what happens after the assessment needs to be set, that is, how the results are distributed and read.

The main results that the solution needs to achieve are therefore :

- To give a quick and relatively accurate result, with a low complexity assessment
- Can the students finish the assessment in a reasonable time?
- To minimize test anxiety and foreign language anxiety. The nature of the assessment can cause fatigue by the time the end is reached, and cause later items to be hastily and incorrectly responded.
- The chosen evaluation method needs to be easily incorporated into an existing web environment

The chosen evaluation method therefore needs to meet all of these criteria, and not have a high complexity of implementation. An administrator needs to have an overlook over the proceedings of each testing/assessment session. Of course, this includes the result gathering. One more important aspect is therefore the result gathering method and result saving method. The gathering of results needs to be easily readable and needs to be accessible by the administrator. The results will depict a relatively accurate level of understanding a certain foreign language. The results should as well be

divided into sessions of results, or groups, because each result set will belong to a certain classroom of students, who will, presumably, do the assessment at the same time.

What is a “test”? A test is a set of instructions, means, techniques and procedures, that lets us determine one’s performance and knowledge. A test can have many forms, it can be multiple-choice, it can be verbal or written, it can be something as simple as a questionnaire. The test must be a mean to measure something, it’s usually knowledge or skill in a certain topic. How is the effectiveness of a test measured? And how do we know that it accurately measures what we want It to measure? The most complex criteria for an effective assessment is validity. For example, a test for assessing the readers ability to read, does not measure previous knowledge, the readers vision, nor something else that is irrelevant to the ability to read. It only measures the ability to read.

The best way to validate the assessment strategy, is to test the tests performance with an interview and assessment of students with known language skills. This will be done in the *Solution* part of this thesis.

Traditional Assessment	Alternative Assessment
One-shot, standardized exams	Continuous long-term assessment
Timed, multiple-choice format	Untimed, free-response format
Decontextualized test items	Contextualized communicative tasks
Scores suffice for feedback	Individualized feedback and washback
Norm-referenced scores	Criterion-referenced scores
Focus on the “right” answer	Open-ended, creative answers
Summative	Formative
Oriented to product	Oriented to process
Non-interactive performance	Interactive performance
Fosters extrinsic motivation	Fosters intrinsic motivation

Table 2. Traditional and alternative assessment [11]

Table 2 shows the differences between two approaches to assessment testing. The traditional method has standardized forms of testing, where the scores of a test alone are sufficient for feedback. The right answers are the only thing that matters, and this is perfect for out test, where less interaction with the test takers is required. It is

important to note that considerably larger budgets and administration for scoring are required to have an alternative assessment in higher education. This is due to the more subjective evaluation and more processing of the feedback given. Of course, this will give a more useful feedback for students, but in this case, that is taking away more than it offers [11].

Tests need to be practical and not expensive in terms of funds and time. Expensive tests that require one on one mentoring are rarely the best solution. In our case, we need one person to be able to administer upwards of a hundred test takers. Practicality of a test is defined with the details of the test being clearly defined and the ability of the students to finish the test in the given time frame. Practicality also depends on the ability of the teachers to administer the test, in terms of equipment, procedural errors, preparation time, budget and time cost for the administration and evaluation feasibility [11].

One of the key issues when designing a language proficiency test, is how the constructs of the language are determined. The language used to construct the very test, must essentially be normal, every day, English. To construct this, with research and sampling, its almost always time consuming, and teachers are almost always consulted to rather use a commercially available test, than construct a new one on their own. More on this will be written in the next chapter.

Multiple choice tests are selective or receptive tests where the test takers select the answers out of a predefined given array of answers, rather than coming up with an answer on their own. The answer options usually range in between two and five given answers per question. Usually, one of the given answers is the correct one, but the test can be formed to have more than one true, or “key” answer(s). The other answers that are not true, are called distractors. Their only purpose, is therefore, to create doubt and, eventually, test the test takers knowledge in the given subject. Distractors and key(s) combined accordingly test the comprehension of the test taker. Distractors, instead of only serving as efficient lures, can also serve as unintentional or intentional clues to the right answer. Distractor efficiency is an important measure of multiple choice tests. The efficiency of a distractor, is its ability to lure test takers, notably the less prepared ones. Taking all of this into account, distractors are going to play a role in the online assessment tool [11].

In regards to grading and student evaluation, all components of grading should be consistent with institutional regulations. Many institutions offer different courses

with different grading criteria. For example, educational institutions in the US use grading in letters A to F, whereas in Germany numbers from 1 to 5 are used, with 1 being the best grade. In contrast to that, in the Balkans, numbers from 1 to 5 are used as well, but 5 being the highest and best grade, whereas a 1 means the test or course was failed. Language assessment is conformed to an almost universal standard around the world, with only minor differences over platforms and countries, contrary to other subjects and fields, which differ greatly. Grades and grading often are done on the go, and tests don't always yield the expected results nor the level of difficulty.

If we specify beforehand standards for grading the performance of students with a numerical grading scheme or point system, we are using an absolute system for grading. Many absolute grading systems follow a model where differentials of 10 percentage points are specified. The key to make such an absolute grading system work is to be very clear on objectives, that is, tasks, tests and various other parts of the assessment, that go into the final grade assignment. One other way of grading is relative grading, which is also more commonly used than absolute grading. This has one major advantage, and that is that the interpreter can adjust or interpret different cases. A distribution of grades is usually within the 100 point or 100 percent area. Such a distribution to 5 levels of grades is very common, and can show us the performance of students very well and also if the test was too difficult or too easy. Most teachers usually bring their own interpretation into the mix, with an estimated distribution, and make minor adjustments to compensate some added difficulty, for example. This attitude towards a relative grading system is very accepted, but often misleading, as teachers' standards of grading often do not match the actual practice [11].

To sum up, grading is usually not based on a universal scale, and is often subjective, and it reflects the institutional expectation for grading. Tests do not always give the desired results, but their goals should be well defined. Across different cultures, grading variations are vast, and need to be understood. In the next chapter, we will look at one of the best known grading and evaluation frameworks that are currently used.

2.2 Existing Language Assessment Frameworks

In order to closer understand the assessment of learning and testing the achievements of foreign language learners, we need a closer look at one of the most used guidelines for learning and teaching foreign languages in the world, the Common European Framework of Reference for Languages, or short, the CEFR.

The CEFR is an European guideline that is used to describe the achievements of people who are learning a foreign language. It was made for the project “Language Learning for European Citizenship” in the 90s. The idea behind it is to officially make a guideline for learning, teaching and assessing for the European languages. As it got increasingly popular during the years, nations around the globe adapted it to many languages that do not originate from Europe. It has since made a great effort in promoting methods of modern language teaching in the world. A need for better language learning and teaching arose from the needs to further the mobility, trade, effective communication and also better access to information, between European states. Language learning is promoted through all educational systems, from pre-school to adult education as well [12].

The uses of the framework range from planning and teaching to certification and assessment. The planning and teaching framework explain different programs, their content and objectives. The certification and assessment frameworks mention the contents of different examinations and go into the assessment details, taking into account positive achievements before negative ones. Basically, some of the best use-cases for this framework include making learning objectives in relation to performance and also self-assessment. Another major use case is with the specification of the content needed for tests and for specifying one’s language proficiency through tests and assessment for later comparison. Taking all of this into mind, the CEFR is the perfect platform to base the test design and evaluation method of this thesis on.

In order to facilitate and describe different levels of language knowledge and proficiency, and help compare different qualifications, the Descriptive Scheme and the

Common Reference Levels have been created in the CEFR. There were many issues faced when designing the reference levels. The first issue was that the scale should be free of context. It should be designed specifically for a purpose for, lets say, a certain job, or a certain school setting. The framework scale should be translatable into all spheres of life of the general population. Such a scale also needs to be simple and user friendly in the context of being accessible and well-rounded understandable in every day life. Assessment measurement issues also played a big part in the designing of the scale. The scale should be objective and precise to avoid error in the very process of measurement. And last but not least, the number of levels must adequately show progression, and avoid being too large, because people must be capable to rationally see a difference and make a distinction between the levels [12].

A scaling framework of six levels was designed that gives an representative cover of the language knowledge relevant to European languages. The division into three levels A, B and C was also made [12].

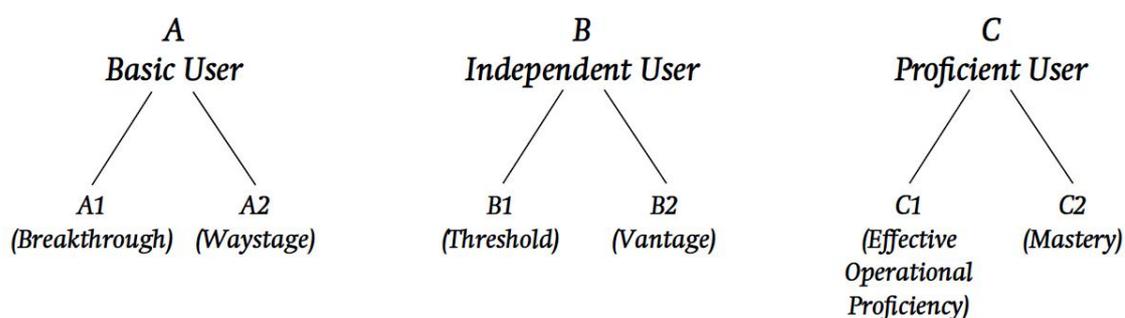


Figure 1. *The layout of the Common Reference Levels* [12]

Basically, this is a division into six levels of achievement. It is thought that these six levels are just enough to cover the actual knowledge differences, but not make a too large, or too narrow gap in between them. These levels don't limit educational institutions in any way in the process choosing how they will organize their own levels and rules. This lead to many institutions having their own versions of the CEFR levels, but all in all, a big uniforming of the assessment practices can be seen in the world, and especially in Europe, after the implementation of the CEFR levels.

The three basic categorizations of the reference levels are into Basic Users (A), Independent Users (B), and Proficient Users (C). The basic user (A1 and A2 levels) can understand sentences and phrases that are frequently used, such as personal information, information about local geography, shopping and similar. This kind of user can communicate in simple terms that allow him the basic everyday communication. On the

other hand, Independent Users (B2 level) can understand more complex ideas and text. This kind of user is able to express himself in the given language in a fluent and concrete manner, even in a conversation with a native speaker without much difficulty. As for the Proficient User (C1 level), this user can use the language for the most complex usages, such as academic and professional. The C2 level is reserved for those who can understand everything read and spoken, even there are different shades of meaning involved [12].

In regards to this thesis, where the main emphasis is on the following up on lectures in an educational institution, we need to take a look at how users from different levels comprehend content and what their vocabulary are.

The levels A1 and A2 are users who can comprehend the usual spoken language, but the emphasis is not set on reading and comprehension of written text. Users at these levels will have a hard time to follow up on a university level lecture in the chosen foreign language. Users at the levels B1 have the ability to maintain interaction, but only to a certain extent, for example, they can follow up on the main point of a conversation, but it may sometimes get more difficult for them to follow into more details. The level next to that, B2, is considered by many institutions the minimum threshold level to follow up on lectures. Users at the B2 levels can comprehend text and language, the relevant explanations, arguments and comments as well. They are able to explain their own view point in the matter, speculate about causes and about different situations. The C1 level and above, are classified as being able to understand fluent and spontaneous language in all forms. This kind of user is almost certain to have no language problems in an university setting [12].

Many higher education institutions in Europe have chosen the B2 levels to be the minimum level to enroll into classes in the native language, but we are seeing a trend in Europe of that level being brought up to the C1 CEFR Reference Level lately. For this thesis, the online language assessment tool will determine as accurately as possible the median level of proficiency of the users, but will leave the final decision to the teacher, to decide how he or she will proceed with the future classes in regards to the results [12].

The online language knowledge assessment tool, needs to give results in accordance to the CEFR Reference Levels, with most students in a reasonable time in regards to the lecture time frame. Given everything said about knowledge levels and how they are defined, it is reasonable to make this test as short as possible. The result gathering method should be quick and efficient as well.

Since the introduction of the CEFR Reference Levels, there has been many calls for integrating them into all levels of education across Europe. Designing tests for learners who are learning their second language, or for short, L2 learners, must take into account different understandings of writing exams. This is a complex and changing process, which include question formation, tenses, different aspects of a language etc. One of the major questions that tells us how well an assessment is designed, is the degree to which it conforms to professional guidelines. For example, the EALTA guidelines used for good practice in language testing and assessment, these guidelines, and many more, have survived over the years to provide reliable testing solutions and ensure the tests are viable [13].

In Second Language Acquisition research, or short SLA research, the performance of someone has been determined and collected in natural conditions, such as real language use. This research has influenced and produced many studies about tasks and task complexity, on L2 learners development. A task is defined as “an activity which requires learners to use language, with emphasis on meaning, to attain an objective”. When doing a task, the user needs to use different aspects of the language to finish it. As far as language specific tasks go, the development of language testing practices and research in this field, has been a major focus in recent decades. Different types of tasks, of course, are used to define different aspects of a language. Language testing therefore has, been looking at L2 learners’ proficiency as multidimensional [13].

When it comes to designing tasks for L2 learners, the tasks should give data which can be used to see differences in performance of each learner. This has usually been done, by using tasks of different difficulty (beginner, intermediate, advanced). The basic approaches are, to either ask all learners to do all three types of tasks, or try to guess the learners knowledge and assign tasks accordingly. These tasks should be able to tell us, what the L2 learners are capable to do, with the given language in question. When it comes to writing tasks, that are usually evaluated by humans, by hand, it comes down, in many cases, to the differences in perspective of the reviewer, and their own language knowledge and perception. One way to try and help here, is by giving reviewers examples of how written tasks are evaluated. This however, is rarely the case, and many language writing tasks, are evaluated relatively subjectively [13].

A proficiency scale, like the CEFR levels, allow reviewer to be more certain and have a good benchmark on the skills of the L2 learners. However, placing a learner on

this scale requires reliable and valid testing, and only then can we place a learner on the CEFR scale [13].

Let's take a look at an interesting paper, published by Feast, at the University of South Australia, in 2002. The paper measured the performance of IELTS test scores, and the performance at University. The motivation for the study, was a large influx of international students, and the need for them, as the Australian government lowered its subsidies for the universities, and therefore the universities relied more for funding on international students. One of the requirements for enrollment is knowledge in the English language, measured by the International English Language Testing System (IELTS). IELTS was designed to "assess the language ability of candidates who need to study or work where English is used for communication". These tests are also required by law in Australia, to obtain a residence permit. The minimum requirement score is 6.0 for enrollment at the universities. Some prior studies, found no significant correlation between IELTS scores, and university performance. They also found, that most of the students who failed at university, actually had quite high IELTS scores. But most of these studies did not have many students with IELTS scores under 6.0, or did not have many international students taking part [14].

The study by Feast in 2002 took therefore the path to answer conclusively, if there is a correlation between IELTS scores and university performance. Another question the study wanted to answer was, if the current minimum IELTS score of 6.0 should be increased, with students having a reasonable chance of performing well later on. The study found, that postgraduate students have a greater chance of success than undergraduate ones, with the same IELTS scores. Science students tend to rely less on their English skills, as their studies proceed, and therefore their IELTS scores become less relevant. Interestingly, most students from Vietnam, whose IELTS scores were lower, tend to give lower university performance over time. Chinese students however, tend to be more successful than non-Chinese students, with the same IELTS scores. All in all, there generally was a positive correlation between IELTS scores and university performance [14].

The 2002 study by Feast also concluded, that the IELTS required score of 6.0 be kept the way it is, because the loss of funding from international students would be too great. However, there have been many reports in Australia, by the teaching staff, that many international students struggle with their English, and questions posed to raise the IELTS admission requirements. The study concluded, that a better choice would be to raise support for the students who already have gotten entry at the current score

requirements. This way, the students will have options to increase their English language knowledge and proficiency skills, without the trade-off of losing many international students before they even apply [14].

Cambridge University offers its language examinations to 4 million people annually, in 130 different countries, based on info from 2011. The first Cambridge Certificate of Proficiency in English (CPE), was issued in the year 1913. It was influenced by the Grammar-Translation Methods, introduced by Meidinger in 1783, and Fick in 1793, which used translation tasks and questions about the use of grammar. The modern methods of language examinations were not in use until the start of the first World War. In the early stages of modern testing adaptations, there was much focus on linguistic forms in the USA, and a structural approach in Britain. Around the 1970s there was a shift from the structural approach, to a more communicative approach, that centered itself on using language as a means of communication. This broadened the evaluations from scholars, to people from all aspects of life. For comparison, the first IELTS test was conducted in 1989. Cambridge University started issuing Business English Certificates in 1993, and this can be indicative of the large influx of foreign businesses, that entered the English speaking markets [15].

From the 1970s onward, there's been much progress to build an European community, and precise language goals were set. The need for applied linguistics increased, and with it, the need for communication skills in the English language. From 1980 and onwards, Cambridge English language examinations became prevalent in the CEFR Reference Levels assessments [15].

When it comes to speaking tests, Cambridge has offered tests of the English speaking ability, with quite a few competences, such as phonology, pragmatics and grammatical knowledge). This is usually tested in a social interaction with instructors in person, and the person who is taking the test. Language that is spoken is evaluated in its social interaction, its ideas and aim to have purpose and be oriented towards a goal. In Cambridge speaking tests, the candidates are often paired to allow different samples of interaction, which is very hard to mimic using technology and without a person who is reviewing, present. Therefore, testing of spoken language is very complex, and there have been many performance assessment frameworks for the spoken language. One of these is the Milanovic and Saville (1996), which gives useful variables for assessing spoken language [16]. This framework was very influential in the rise of the Cambridge ESOL examinations in the 1990s. These examinations included face to face speaking assessments, in which examiners are also speaking, asking questions and having

involvement. Over the years, this has sparked questions as to, how much space do test takers have to talk, how much does this affect their abilities and hinder them, or cause anxiety [17].

Taking all of this insight into account, it is clear that the future for technology based examinations is very bright, and should be explored and developed further. In the next chapter, the existing technologies used that are currently widely used in online language assessment, will be examined.

2.3 Dialang - Language Assessment Framework

There are many online language knowledge assessment tools available, especially in the English language department. Some stick out more than others, many can be found that don't conform to any set standards or frameworks. One of the best known ones, that uses the Common European Framework of Reference Levels as a basis for determining language proficiency, is called Dialang.

The Dialang tool was made by the University of Lancaster in the United Kingdom and was funded by the European Union. The goal was to make an online tool that would test the language performance of subjects, and that would conform the the Common European Framework of Reference. Dialang is used to evaluate reading, writing, listening and the vocabulary of a person, in 14 different languages. Important to note is, that Dialang is not made for a specific language or course, but its based solely on the language specifications suggested by the CEFR [20].

There are numerous advantages to an online based approach to language learning and assessment. One of them is the advantage of real time scoring of the subject's tests. The widespread availability of computers has also lead to general acceptance of computer-based testing and learning platforms. Flexibility in time, ease of writing the tests and immediate feedback, inexpensive tests, these are all perks of online-based testing. However, some draw backs exist, such as incompatibilities, server failures and cheating. In the case of Dialang, according to Alvarez and Rice in 2001, the advantages are: objective scores, immediate feedback, language level identification, it offers study opportunities, results can be stored to match future results [21]. Alderson in 2005 also mentioned that one of the main values of Dialang is that the system is free of charge [22]. Areas where Dialang has had disadvantages are limitations in certain languages and the fairly traditional theme in which the test is set. Alvarez and Rice argued that Dialang lacks the possibility to actually assess real productive skills, such as the compiling of sentences, which a real person could assess [20].

In 2014 at the University of Tehran in Iran, a study has been conducted to diagnose a language learners skills in a self-assessment setting, based on Dialang. The main question to answer, was is there any statistically significant variability in the learners performance in the listening, reading and writing sections. The study found that

even after a six year English language education at the secondary level and three years at the post-secondary level, students did not report high scores. Researches argued, that, before the university level studies, most education is less focused on writing and speaking skills, and more on reading skills. A big factor contributing to the low scores is also the more traditional educational approach, rather than an institutional framework focused one. Researches concluded that Dialang is better suited for the self assessment of university goers [20].

Since Dialang was first put into use, its usage and research was mainly focused on adult users. A research conducted by Baglantzi in 2012 in the Greek Junior High School context tried to answer the question if Dialang could be used in a high school. Can high school students benefit from the feedback and procedures of Dialang? And is Dialang actually helpfull to the teaching staff. The research was conducted on high school students with a similar curricula to other member states of Europe. The study involved a questionnaire before and after the students' use of Dialang. In Table 4 we can see how the Dialang test placed the students according to different parts of the testing procedure, and most of the students are in A1 for reading, writing and listening. If we look at the students' self-assessment results in Table 5, and the teacher's assessment at Table 6, we can see a correlation in the listening section. If we however take a look at reading and writing, we will find that the teacher agrees more with Dialang when it comes to the reading skill, and the students agree more with Dialang when it comes to the writing skill [23].

Dialang Test	Levels			
	A1	A2	B1	B2
1. Listening	12	4	3	1
2. Writing	8	10	1	1
3. Reading	8	7	4	1
4. Vocabulary	6	7	6	1
5. Structures	6	4	6	4

Table 3. Frequency of students across levels according to Dialang test results [23]

Levels

Dialang Test	A1	A2	B1	B2
1. Listening	6	6	8	-
2. Writing	1	10	4	5
3. Reading	-	9	10	1

Table 4. Frequency of students across levels according to student self-assessment [23]

Dialang Test	Levels			
	A1	A2	B1	B2
4. Listening	5	8	7	-
5. Writing	7	7	6	-
6. Reading	4	8	7	1

Table 5. Frequency of students across levels according to teacher assessment [23]

The study also found that Dialang is very practical because of its automation. The provision of feedback and scores is done automatically and all the teacher needs to prepare is the form on which students report their feedback. The results of the study show that high school students consistently overestimated their language abilities. Students reported that testing their skills “unofficially” via Dialang, helped estimate real life situations and spot their weaknesses [23].

To be considered for educational and placement purposes, Dialang must first be available on the institutions computers, so that a class of students, can take part at a time. This means the software needs to be either downloaded and installed, or it must be used online, as is. In order to shorten the time required for the taking of the test, proposed have been a number of introductory methods for the students [23].

One thing that was surprising in the study by Baglantzi (2012) is the findings about what feedback students value the most. Only a small percentage of the students actually reviewed their questions, to check in details what answers they got wrong, and what they got right. Students showed a leaning towards a numerical evaluation scale, where they don’t need to explicitly understand what their score on that scale means. This goes to show that teachers usually don’t explain in detail, what the scores of a grade actually represent, and most importantly, how to achieve that better score and for what effort [23].

Students that took the Dialang test, reported that they did not know, to an extent, what the CEFR levels actually represent. They also reported that they were placed, in previous exams and evaluations, in a certain CEFR Reference level, without being explained what that level is [23].

A study from Ajmal from the International Islamic University in Islamabad, Pakistan, was conducted about the use and interface between learning and assessment of Dialang in 2020. The study was based, in this case, on undergraduate students, as opposed to the study of Baglantzi in 2012. [17]

The problems that undergraduate students mostly report when it comes to the English language, is vocabulary and comprehension. These problems are especially present in the social studies and science vocabularies. Different ways to learn new languages have drastically evolved since the 1970s. “Strategic Learning” has become different from standard learning. For example, instead of using standard learning to look into a vocabulary of words to memorize them, strategic learning has developed word repetition strategies, finding similar words, underlining, overlearning, and many more new methods [24].

Gender is also a factor when it comes to learning strategies. Women are found to use more learning strategies than men. The use of different learning strategies also is different geographically. It was found that women in the western culture use different learning strategies than those in eastern cultures. Of course, the learning process is not only dependant on the learning strategy, but on many other factors as well [24].

The research done by Ajmal in 2020, was focused on one group of students from the University of Lahore, who were participants of an English language course. It is important to note, that the English language course was designed to conform to the Economics courses, and therefore was expanded to include many idioms and words used in Business and Economics. The research focused on their learning strategies and the evaluation of their English language knowledge. In the year 2018/19, the students completed the Dialang Test in all the fields, which include reading, writing, listening, grammar and vocabulary. The students used 30 different learning strategies, and their knowledge was evaluated at the end of the academic year again, using Dialang and Oxfords ESL/EFL online test [23].

The majority of students received a B2 Dialang vocabulary score, 28% got C1 or higher. As far as the grammar test, 44% got a B2, which was better than the vocabulary test, and 38% of students got a C1 or higher in vocabulary. The purpose of the study, was to, with the help of Dialang, show how learning strategies can help students gain a language proficiency. In the Dialang vocabulary test, social and memory strategies helped gaining a higher score. Social strategies, which include speaking the language with native speakers, helped gain an overall better score in Dialang. Making relations with what a student knows, and some new knowledge in the language, helped gain an overall better score as well. An important finding was, that many students who gained good scores in Dialang, reported watching TV shows in the English language to help them connect their knowledge. This strategy can be considered a cognitive language learning strategy [23].

All in all, the study of Ajmal in 2020, did not only show that using different learning strategies positively resulted in student gaining proficiency in given language. It also showed that Dialang, has proven to give consistent results, with far less effort for the teaching staff, than other more traditional knowledge assessment methods. It is important to motivate students to use different learning strategies, but also the teaching staff in higher education facilities, to use alternative knowledge assessment methods. This is especially prominent in less developed countries, where the use of online tools has not gained such momentum yet. Tools like Dialang should be implemented not only in universities, but at all levels of education [23].

All in all, Dialang has proved as very good in bringing reliable results to students in a high school setting, as well as an university setting. The next chapter will draw a line under the research done, and a solution for the online language knowledge assessment tool can be made.

2.4 Research Summary

In the research conducted in this thesis, an introduction to the problem was made, as well as the importance and significance of online testing and knowledge assessment. A similar technology has been described, called Dialang, that is used worldwide for the purpose of online language knowledge assessment.

To sum up the importance of Dialang, Dialang is a very useful tool, recognized and used by many institutions around the globe, and should be a major reference point for this thesis. As we can see from the research of Baglantzi in 2012, Dialang has been found to be very practical and also helped students in the evaluation procedure of their knowledge. Therefore, Dialang should be used as a classic example of why and how this kind of knowledge assessment can and should be used.

Testing in general, should be approached with regards to the student's well being. Many factors, as described in this thesis, lead to the development of tools and ways for knowledge assessment. These include anxiety, language skills, cultural differences, ethics, rules and regulations. When designing a quiz, or any kind of knowledge assessment, these factors must be taken into account, to ensure that all students have an equal chance, and the same starting grounds.

With all of this in mind, the solution will be made to best assess a student's CEFR Reference Level of the English language, and show the data to the web platform administrator.

In the next chapter, the Solution for the language assessment tool is described.

3 Solution

3.1 Technology used

One of the challenges faced is, that many lecturers and professors have an existing web page, dedicated to the studies that they are working on. This thesis is taking an approach in such a way, that the given solution will be easily incorporated onto existing web pages. This will be done by using the existing technology called Drupal.

Drupal is an open source web content management system, which makes it possible to create web content and web sites in a quick and easy way. A content management system, or *CMS* is an application that is used to manage the customization and creation of web content. A CMS can support one or more users in the form of administrators. In this case, we will look at a website as the reference. The contents of a website can range from simple to complex, and they are subject to change. The basic idea behind a CMS is, that users with little to no expertise in programming and the actual backbone of a certain application, can manage its content with less effort as opposed to learning how the code behind it works. This is achieved through a front-end user interface that lets a certain user modify the contents and update them, without having to deal with the actual back-end code. Time saving and security are also advantages of content management systems. A CMS can save much time, because it helps us in managing a wide variety of different types of content. The use of content management systems has greatly risen in the past decade, and web administrators have the ability to publish content and change the appearance of their content, without advanced programming knowledge [18].

Taking all of this into account, it makes sense to incorporate the solution into an online tool, and into an existing CMS, Drupal. Drupal is used to make websites, and it is modular. It has basic functionality, but functionality is added with enabling either

modules that are built in, or the ones that are from a third party. So basically, adding modules is customizing Drupal. The Drupal version that is used in this thesis, is going to be *Drupal 7*.

Language	PHP Database Abstraction Layer
Database	MySQL / PostgreSQL / SQLite / ...
Web Server	Apache / lighttpd / ISS / ...
Operating System	Linux / BSD / Mac OSX / Windows / Solaris / ...

Table 6. *Drupal's technology stack* [18]

Drupal runs on all operating systems that support PHP, such as Linux and Windows. As for the web server that is mostly used with Drupal, its the Apache Web Server. In the database spectrum, Drupal supports all databases that support PHP. The most widely used database with drupal are MySQL and PostgreSQL, as can be seen in *Table 3*. The language used and in which Drupal is written in, is PHP. Drupal comes with a core, which includes basic functionality and blocks for websites. The administrative interface has easily accessible administrative functions in a menu that is at the top of the page [18].

Phase	Purpose
Configuration	Set global variables used throughout the bootstrap process.
Database	Initializes the database system and registers autoload functions.
Variables	Loads system and all enabled bootstrap modules.
Session	Initializes session handling.
Page Header	Invokes hook_boot(), initializes the locking systems, and sends the default HTTP headers.
Language	Initializes all the defines language types.
Full	The final phase: Drupal is fully loaded by now. This phase validates and fixes the input data.

Table 7. *Drupal Bootstrap phases* [18]

One important thing to note is that Drupal bootstraps itself on all requests while going through a series of phases, as can be seen in *Table 3* [19].

Modules are the basic building blocks of Drupal, and they carry functionality, which can be enabled when the module is installed. So, the solution of the language knowledge assessment, will be incorporated in such a module, which can then be added onto existing Drupal websites. We can differentiate between two main types of modules, core and contributed. Core modules are the ones that come with the Drupal core, some of them include modules for menus, search bars, and for example, forums. Contributed modules, are however, third-party sourced, but they still have to follow rules and guidelines. Modules are, in essence, what makes Drupal modular, and this creates the emphasis on collaboration and customization. Modules are what enables us to create internet portals, online blogs and newspapers, business websites, e-commerce websites, image galleries, and much more, all using Drupal and its modules [19].

As said, modules that are not Core modules, are third-party sourced. They can be installed by users on their own Drupal adaptations to use the Modules' functions. So for example, we can install a module that extends Drupal's functionality by adding new content types, such as blogs, different files, notifications, peer to peer publishing and much more [19].



Figure 2. *A new Drupal Website Dashboard* [19]

For this solution the Drupal 7 version of the Drupal Core is used. The web server that will be used is the *Apache* server, with a *MySQL* database, running on *Windows 10*

operating system. The language used is, PHP. The solution will be a Module for testing and assessment of a students' language skills. The module will have the options for adding and removing different parts of the quiz, checking the results and some other functions that are explained in detail in the segment *Solution*.

3.2 Solution

The solution of this thesis is based on the research done and documented in the thesis. An online language knowledge assessment tool will be made as part of a Drupal 7 Module for expanding the use and functionality of an existing Drupal based web platform.

An English language Quiz will be made, that students can use, on their own devices, online. The results will be stored, for the administrator to see, and see how well students performed. The results therefore can be used to plan the lectures in accordance to the students' language knowledge levels.

The Quiz, will be designed to last at most, 5 to 10 minutes, for the sake of the emphasized usage, before a lecture has begun. In this time, the best way to assess one's language knowledge, is by making the answers predetermined. Therefore, the result gathering method will be automatic, and not require a professor to check the answers. The quiz feedback will also be automatic and instantaneous. All the answers and results will be stored online, for the students and the administrator, the professor in this case, to see.

The quiz will be made up from 20 predetermined questions, where either one or several of the given answers are correct. The type of the questions will be multiple-choice, so that the system can grade the result automatically and instantly without a reviewer having to check the results and grade them manually. The pool of questions will be stored online as well. For the functions and flow of the test, the existing Drupal 7 module, called Quiz [25] , will be used. The Quiz module has the functionality of creating graded assessment tests in Drupal, or a series of questions. The answers are stored in the database of the web platform. The questions will be designed, so that they conform to the CEFR Reference Levels.



Figure 3. *The Drupal Quiz Module with a new Quiz created [25]*

In Figure 3, we can see a newly created quiz, using the Drupal Quiz module. The Quiz can then be configured to contain various types of questions. There are also many options to configure the grading and result display options.

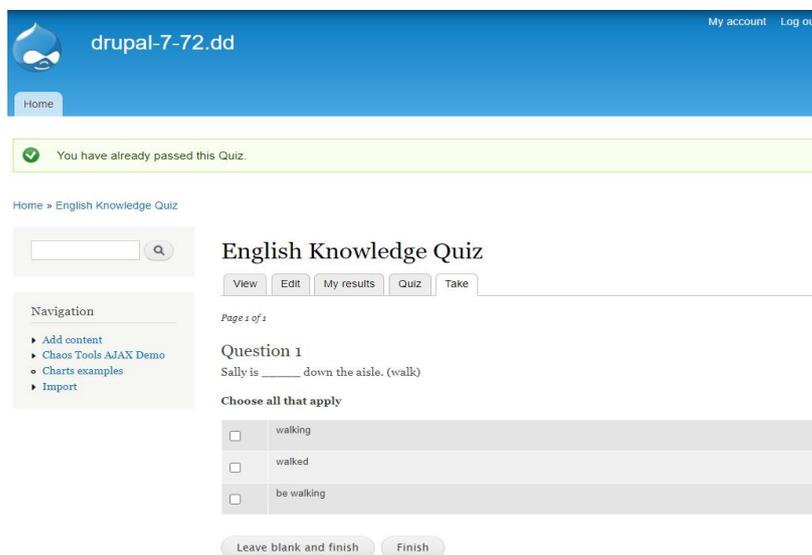


Figure 4. *The Drupal Quiz Module with a quiz that has been started [25]*

In Figure 4, we can see a quiz that has been started and a question with in the multiple-choice mode. We can predetermine the right answers, so that the system can grade the quiz automatically, and instantly show the results to the user.

The results will be graded in accordance to the question answers. This quiz will grade users up to the B2 level, which is currently the official requirement level for most English taught study programs in Austria.

Score	Grade
0-24%	A1
25-49%	A2
50-74%	B1
75-100%	B2

Table 7. *Result-score matching in the solution Quiz*

As we can see in Table 7, each CEFR Reference level in question (A1, A2, B1, B2) has its own range of the score.

Number of questions	Points for each question
5	1
5	1
5	1
5	1

Table 8. *Assigning points to each of the 20 questions*

As we can see in Table 8, out of the 20 questions in the question pool, each question will carry 1 point. Because there are 5 questions for each level of difficulty (A1, A2, B1, B2), this standard grading will ensure that only if all, or most of the questions of the lower tier levels are done correctly, will get a users point up to reach a score that gives a higher CEFR level as a result. For example, if a student does all questions from the A1 and B2 lineup correctly (total of 10 points, 50%) , the grade at the end would be B2.

The next chapter will go over how the questions were designed and formatted.

3.3 Question Design

For the design of the question, a thing to keep in mind in the design process should be, that they should conform to the adequate CEFR Reference Level difficulty. Also, the theme for the questions, will be computer science and project management.

All the questions, for all separate CEFR Reference Levels, were created to bear the form and resemblance to the official Cambridge English assessment tests of the according CEFR levels (A1[27], A2[28], B1[29], B2[30]).

The only thing that was changed, was the theme of the questions, from general, to computer science and project management. In the next table, the multiple-choice questions are laid out, and the correct answers are marked in bold.

ID	Question	Answer
A1_1	"Where do you study?" I study _____ Vienna.	in ;into;at;on
A1_2	_____ five laptops on the table.	There are ;Their is;There is;Their are
A1_3	What is Anna doing now?	She's in a project meeting ;She is a project meeting.;She's a project meeting.
A1_4	I _____ talk on Slack now, because I'm driving.	can't ;don't;not;no
A1_5	I have a programming tutor. _____ name's Annabela.	Her ;He;She;His

Table 9. *Questions for the A1 CEFR Level*

In Table 9 we can see the questions for the A1 level, these are the simplest questions in the Quiz.

ID	Question	Answer
A2_1	Coding in Java is _____ than coding in Python.	easier ;easier;easier;easier
A2_2	Adam is _____ many fields and is a _____ - known mathematician.	interested in/well ; interested on/well; interested in/good; interesting on/ good
A2_3	What are his past experiences?	He has been to a physics competition. ; He has to a physics competition been.; He has went to a physics competition.
A2_4	They _____ when the phone rang.	were in a meeting ;were on a meeting; have a meeting;are meeting
A2_5	Would you _____ that we didn't go to the lecture?	rather ;before;like;love

Table 10. *Questions for the A2 CEFR Level*

ID	Question	Answer
B1_1	We _____ to university, but now we take the train.	used to drive ;use to drive;used to drove;use to drove
B1_2	I love to code, but _____ coding since 2010.	I haven't been ;I didn't went;I haven't went;I didn't been
B1_3	Their project leader_____ their weekend, before they spoke about it.	had already planned ; has already planned; did already plan;is already planning
B1_4	You were late for an important meeting. You _____ home earlier.	should have left ;must have left;should had left;must had left
B1_5	"Sarah had an math exam yesterday and she's really nervous about the results." Choose a similar adjective for "nervous"	worried ;confident;positive

Table 11. *Questions for the B1 CEFR Level*

In Tables 10 and 11 we can see the questions for the A2 and B2 levels, these questions are still relatively simple in regards to a university level English requirement. As previously stated, most universities, where English is required, require a B2 CEFR level of language knowledge, or above.

ID	Question	Answer
B2_1	I don't mind if you upgrade the software, _____ you back up the hard drive first.	as long as ;while;whereas
B2_2	I _____ for the last few months, because I switched to the back-end team and am getting used first.	haven't been working ;'m not working;haven't working;wouldn't have worked
B2_3	I'm sure that _____ her first application by the end of the year.	she'll have finished ; she'll be finished;she has finished
B2_4	"The exam started ten minutes ago, but I couldn't take it because I didn't bring my laptop." If I _____ my laptop, I _____ taking the exam right now.	hadn't forgotten / would be;had remembered / will be;had remembered / wouldn't be;hadn't forgotten / will be
B2_5	It is _____ difficult to ensure the security of user data.	extremely ;totally;absolutely

Table 12. *Questions for the B2 CEFR Level*

In Table 12 we can see the questions for the B2 level, these questions are in the difficulty level that is usually required for attending most Bachelor Degree programs in Europe. All of the questions were designed to conform the described CEFR levels, 5 questions for each of the 4 CEFR levels in question. Therefore the questions should only be changed by an administrator, as to not mix up questions of different difficulties in the questions pool, and thus create an dis balance in the results. This even spread of questions will create an evenly spread result set, to easier see the differences between different knowledge levels.

The questions will be imported onto the new module, using the Drupal 7 Quiz Question Import [26] module. A CSV (comma separated values) File will be created, with the questions inside, and then uploaded onto the web platform using the Quiz Question Import Module.

4 Conclusion

As the influx of international students only increases, and the globalization with the help of the English language broadens, assessing the language knowledge of students, especially when it comes to the English language, but also many others, will be of high importance.

When it comes to the evaluation of the solution tool, different aspects need to be taken into account. One of the first aspects is, how much effort should a reviewer, or professor at a university, put into evaluating this tool.

One of the methods to evaluate this tool, is to, of course, ask students about their perceived language knowledge level before taking the test. Another could be, evaluating different groups of students, as they do English language intensive workshops, homework, or some similar exercise, that is already planned beforehand for the study course. In these kind of exercises, students who get good results in the online tool, should also be the ones who get better grades and scores in the exercises. If this is the case, then the tool is a good approximation of students' language knowledge.

In this thesis, the importance of international expansions has been highlighted. This only furthers the narrative that online testing should not only be encouraged, but also be a major part in every educational institution around the globe. The global pandemic events of 2020, only further this narrative, and pave the way for future generations to have a system, where more effort is put into developing technological solutions, to ease our lives, and make up more time for us to pursue other things.

5 References

- [1] Niall Hegarty, *The Presence and Importance of International Students to Universities in the United States*, Tobin College of Business St. John's University, New York (USA), 2014
- [2] Douglass, J.A., & Edelstein, R. *The global competition for talent: The rapidly changing market for international students and the need for a strategic approach* UC Berkeley, (2009).
- [3] Fee, E. J., Radford, *Promoting student and institutional success in Canadian colleges and universities*, Simon Fraser University, Washington, USA, 2011
- [4] Clayton Smith, *International Student Success*, University of Windsor, Ontario (Canada), 2016
- [5] Horwitz, E.K., M. B. and Cope, *Foreign language classroom anxiety. Modern Language Journal*, 1986
- [6] Mary J. Pitoniak, John W. Young, Maria Martiniello, Teresa C. King, Alyssa Buteux, Mitchell Ginsburgh, *Guidelines for the Assessment of English Language Learners*, Stanford University, USA, 2009
- [7] Matsuda, Gobel, *Anxiety and predictors of performance in the foreign language classroom*, Kyoto Sangyo University, Japan, 2003
- [8] Phillips, E., *The effects of language anxiety on students' oral test performance and attitudes. Modern Language Journal* 76, 1992
- [9] Masoomah Salehi, Fahimeh Marefat, *The Effects of Foreign Language Anxiety and Test Anxiety on Foreign Language Test Performance*, Allameh Tabatabaee University, Iran, 2014

- [10] Dr. Selami Aydin, *Test Anxiety in Foreign Language Learning*, Istanbul Medeniyet University, Turkey, 2006
- [11] H. Douglas Brown, *Language Assessment Principles and Classroom Practices*, San Francisco State University, USA, 2004
- [12] Council of Europe, *Common European Framework of Reference for Languages: Learning, Teaching, Assessment*, Language Policy Unit Strasbourg
- [13] Riikka Alen, Ari Hunta, Mirja Tarnanen, *Designing and assessing L2 writing tasks across CEFR proficiency levels*, University of Jyväskylä, Finland, 2010
- [14] Vicki Feast, *The Impact of IELTS Scores on Performance at University*, University of South Australia, Australia 2002
- [15] Dr. Hanan Khalifa, Coreen Docherty, *Cambridge English Research Notes Issue 51 February 2013*, University of Cambridge, 2013
- [16] Saville, *The process of test development and revision within UCLES EFL*, in Weir, C J and Milanovic, M (Eds) *Continuity and Innovation: Revising the Cambridge Proficiency in English Examination 1913–2002*, Studies in Language Testing volume 15, Cambridge: UCLES/Cambridge University Press, 2006
- [17] R, Alexander, J Gray, *Issue 13 of Research Notes : The Cambridge Primary Review Surveys*, University of Cambridge, 2012
- [18] Manoj Kumar Srivastav, *Web Content Management*, St. Xavier's College, Kolkata, India, 2016
- [19] Todd Tomlinson John K. VanDyk, *Pro Drupal 7 Development*, New York, USA, 2010
- [20] Taghizadeh, Alavi, Rezaee, *Diagnosing L2 Learners' Language Skills Based on the Use of a Web-Based Assessment Tool Called DIALANG*, University of Tehran, 2014

-
- [21] Sasad Alvarez, M., & Rice, J. *Web-based tests in second/foreign language self-assessment. The 29th Annual Proceedings*, Dallas: The National Convention of the Association for Educational Communications and Technology, 2006
- [22] Alderson, Ch. J., *Diagnosing foreign language proficiency: The interface between learning and assessment*, London, 2005
- [23] Vassiliki Baglantzi, *Online Diagnostic Assessment: Potential and Limitations (The case of DIALANG in the Greek Junior High School Context)*, University of Athens, Greece, 2012
- [24] Muhamma Ajmal, *Using DIALANG in Assessing Foreign Language Proficiency: The Interface between Learning and Assessment*, Islamic University Islamabad, Pakistan, 2020
- [25] Drupal 7 Quiz Module , <https://www.drupal.org/project/quiz>
- [26] Drupal 7 Quiz Question Import Module, https://www.drupal.org/project/qq_import
- [27] Cambridge English Exam A1 Starters Preparation,
<https://www.cambridgeenglish.org/exams-and-tests/starters/preparation/>
- [28] Cambridge English Exam A2 Key Preparation,
<https://www.cambridgeenglish.org/exams-and-tests/key/preparation/>
- [29] Cambridge English Exam B1 Preliminary Preparation,
<https://www.cambridgeenglish.org/exams-and-tests/preliminary/preparation/>
- [30] Cambridge English Exam B2 First Preparation,
<https://www.cambridgeenglish.org/exams-and-tests/first/preparation/>

Attachments