

Exploring test design in measuring readability to create evaluation of text simplification software for individuals with ID

Author: Eliaz Brzeskot Ganning, masters student at Linköping University

Mentor: Åsa Elwér, Senior Lecturer at Linköping University

Abstract

This paper seeks out to explore different test designs of a reader comprehension test for people with intellectual disability (ID) to better understand reading processes and what future text simplification software should focus on. A theoretical background in cognitive science is presented, with theories of mental processes involved in reading comprehension. The paper consists of two pilot studies with a total of three high school students with ID that each completed a questionnaire with lexical- and sentence-oriented questions. Both quantitative data from the questionnaire answers and qualitative data from reflection, observation and interviewing was collected. Based on these data, the questionnaire was modified after the first pilot and discussed after both pilots. The results and discussion do not conclude any fixed solutions or complete test design, but presents less effective design options and underline problems with meta cognitive reflection in tests on individuals with ID. Further scientific exploration in the area is required.

Keywords: Intellectual disability (ID) – reading comprehension – text simplification – meta cognitive reflection

Acknowledgments

Special thanks to the high school that agreed to participate in the study, especially all the very helpful teachers and inspiring students. I would also like to thank Linköping university and my mentor Åsa Elwér for continuous feedback and help.

Introduction

Reading is an essential tool for both learning during education and understanding information in everyday life. Although our digital society provides means of receiving information in forms of auditive and visual media (video and animation), the information presented in text seems to be nothing we can completely substitute, and maybe never will. Reading and understanding the printed word is therefore still an important part of life. To fully comprehend a text, it is suggested that a person should be able to understand 95% of the content words of a text (Paris, 2005). This creates problems for people with low word knowledge, which often have disabilities affecting reading. Struggling readers are prone to miss everyday information presented digitally through smartphones or presented physically in public areas and tend to develop frustration and anxiety in connection to reading (Dudley, 2005). In an effort of ensuring reading comprehension for the heterogeneous population there has been a growing interest in *test simplification* software. Due to the varying reading difficulties and abilities, depending on the individual, it is challenging to create a text simplification method applicable for everyone. This project focus on adolescents with *ID (intellectual disability)* and how different types of text adaptation methods can facilitate reading comprehension.

Research in text simplification for individuals with ID is an area few studies has examined to this point. The aim of this study is to create a test that can be used to analyze text simplification principles based on reading behaviors of individuals with ID. This will be based on self-reported comprehension where test subjects to explain what they understand and prefer to strengthen reading comprehension. Such a test, can be used to evaluate automatic text simplification on subjects with ID. Hence, there is the necessity for creating this test and to ensure that it can be understood and performed by adolescents with ID. This paper will discuss some theoretical principles behind reading comprehension, intellectual disability, and text simplification and with a qualitative method explore the design of an ID reading comprehension preference test. The research question is: *how can we examine self-reported comprehension and textual preferences in adolescents with ID?*

Theoretical background

Reading comprehension in cognitive science

The early cognitive and psychological theories about reading comprehension focused mostly on memory processing and how readers internal representations of a text differed from the textual representation of the text (Lorch & Broek, 1997). It was not until the 70's that the scientific area began focusing on integrating theories about decoding and linguistic processes. To fully understand reading comprehension, the process of reading must be considered and not just the derived understanding that occurs first after reading (Broek, Helder, & Karlsson, 2014). Successful reading comprehension emerges as correct cognitive processes are carried out step by step whilst reading. This study reviews the theoretical background of reading comprehension.

An important definition of reading comprehension is expressed as the product of decoding (word identification and phonological understanding) and linguistic comprehension (understanding text content and language structures) (Hoover & Gough, 1990). This concept is called the *Simple View of Reading* and highlights the symbiotic nature of linguistic comprehension and decoding, and how both are required for an overall successful reading comprehension.

Another theory about reading comprehension states that the reading comprehension is almost entirely dictated by the lexical knowledge of words. This theory, called *the lexical quality hypotheses*, argues that individuals with higher quality knowledge of words, (defining the orthography, phonology and semantics of a given word) is more likely going to comprehend a text better than an individual with lower quality knowledge of words (Perfetti & Hart, 2002). Furthermore, Perfetti and Hart argues that the substitution of less frequent words with more frequent words aids reading comprehension for both high skilled and low skilled readers. For text simplification this means that replacing difficult words with simpler synonyms should be sufficient for repairing reading comprehension.

The theories of the simple view of reading and the lexical quality hypothesis are quite specific and does not describe the width of processes present in reading comprehension. In an attempt of including the width of reading theories, Perfetti and Stafura (2014) presented a general *reading systems framework*. This framework is defined as: the knowledge behind

reading ability (linguistic, orthographic, and general knowledge), reading process (decoding and understanding content of sentences) and the connection between these processes and other cognitive systems (such as memory and perception). Reading comprehension difficulties can be identified in specific subparts of the framework and the framework show how these subparts affect the level of comprehension.

Methods of testing reading comprehension often revolves around asking questions about the content of the text, either by asking for literal or inferential content (Arfe, Mason, & Fajardo, 2018). *Literal questions* refer to questions which can be answered by the information explicitly stated in the text. Questions that require deeper contextual understanding and inference to answer, are called *inferential questions*, and are often more difficult for struggling readers to answer. One of many explanations to this is that struggling readers often lack the sufficient prior knowledge to infer implicit information while reading. An example of an inferential question could be to describe a relationship between two characters in a text as good or bad, provided that it is never explicitly mentioned. Hence, one method of simplifying text is not just by replacing difficult words with more common synonyms, but also changing structure and expanding information that is not explicitly being mentioned.

Reading for individuals with ID

The American Association on Intellectual and Development Disabilities defines ID as (Definition of Intellectual Disability, retrieved: 2021):

“Intellectual disability is a disability characterized by significant limitations in both intellectual functioning and in adaptive behavior, which covers many everyday social and practical skills. This disability originates before the age of 22.”

ID is recognized by measuring IQ score (often lower than 70-75 IQ) or by identifying difficulties with some cognitive abilities. Reading comprehension skills in individuals with ID varies depending on impaired level of the other literacy abilities. People with ID seems to be going through the same development in reading as typical children do at a low age (Wingerden, Segers, Balkom, & Verhoeven, 2016). However, they tend to develop in decoding at a slower rate compared to typical readers (Wingerden, Segers, Balkom, & Verhoeven, 2014). In some individuals with ID, the decoding skills can be learned to an extent where lower-level reading comprehension does not differ significantly from typical readers. However, the higher-level

reading comprehension, involving the contextualizing of abstract words and sentences, is always impaired in children with ID compared to typical readers. Lower syntactic and morphological ability in individuals with ID, is also more likely to be connected to late development and should be regarded in perspective of the subject's mental age (Koizumi, Saito, & Kojima, 2019).

Delayed development of decoding is one factor that affects reading skills. However, lowered decoding skill is not linked to level of general intelligence, but rather linked to weaker phonological working memory (Connors, Atwell, Rosenquist, & Sligh, 2001). For people with ID, reading becomes dependent on refreshing and rehearsing of phonological strings in the working memory. Thus, impaired working memory affects decoding skills which also affects reading skills. Level of phonological awareness, the ability to manipulate segment of speech, has also been shown to affect reading skills for students with ID (Barker, Sevcik, Morris, & Romski, 2013). In views of the *lexical quality hypothesis* individuals with ID often possess both lower lexical decoding skills and smaller mental lexicon, which could affect reading (Fajardo, o.a., 2014).

Investigating reading comprehension in individuals with ID

Investigating reading comprehension in a specific target group, in this case subjects with ID, include a couple of considerable challenges to the research. One is focusing on the experience of people with ID and studying them in a collaborative manner, instead of studying the subjects (Knox, Mok, & Parmenter, 2000). Since individuals with ID are the target group, they should also be considered the experts of their domain, providing information from a personal perspective. But qualitative research with people with ID is complicated and difficulties can sometimes be attributed to aspects of the individual instead of problems with the methods. Interview based research before the 80s on people with ID tended to attribute errors on the interviewee instead of reevaluating the interview approach (Hollomotz, 2018). Therefore, is it important to focus on an inclusive research method and work with aiding the subject throughout the test. It is also important to create a trustful relation during interviews and not to ask questions that are too complicated or require time pressured responses (Llewellyn, 1995).

This study will investigate reading comprehension by comparing selection of textual alternatives and examine which underlying reasons result in these selections. As mentioned above, these underlying reasons can only be expressed by the person with ID and is best carried out through a qualitative approach. But to understand experienced difficulty in people with ID through qualitative investigation requires the subject to engage in metacognitive and self-reflective reasoning. An intellectual disabled adolescents' metacognitive ability is comparatively on the level of a typical developing child in the ages of seven to nine. With that being said, self-reflection is often affected by factors such as feeling of difficulty, unfamiliarity with a task and external negative emotions (Pennequin, Igier, Pivry, & Gaschet, 2021). Evoking metacognitive reflection may become difficult but is necessary to understand the thought behind a selection. So even though some quantitative data will be collected through the study, the main research approach will be qualitative on semi-structured interviews.

Text simplification

Text simplification is an area within the field of *natural language processing* (NLP) and refers to the automatic or manual processes of altering text content and structure to aid language understanding (Shardlow, 2014). A text can both be simplified in *readability*, how easy the text is to decode, and *understandability*, how easy the text is understood (Rello, Baeza-Yates, Dempere-Marco, & Saggion, 2013). For instance, one study found that there is improved readability for students with dyslexia when a text has high-frequent words and improved understandability when the words generally are shorter. Indeed, this study shows how text simplification can be used as a tool to aid people with reading problems. A similar text simplification study has been carried out with children as target group with somewhat promising results (Belder & Moens, 2010). However, there has not been much research carried out on text simplification focused on individuals with ID, other than studies exploring connectives and possible simplification tools in relation to individuals with ID (Fajardo, Tavares, Ávila, & Ferrer, 2013) (Chen, o.a., 2017).

The difficulty of a text depends, amongst other factors, on text characteristics which is defined by the perceived and actual difficulty (Leroy, Endicott, Kauchak, Mouradi, & Just, 2013). Perceived difficulty is expressed by the words and their part of speech tag, which becomes more difficult if, for instance, a text has many low frequency words, few verb phrases, and verbs. The perceived difficulty can be analyzed by for example, having the participants rate a

word's difficulty on a five-point scale (Leroy & Kauchak, 2014). Actual difficulty depends to the underlying and overarching attributes of a text such as topic, composition, the readers knowledge, and the contained information. This study is testing the perceived difficulty of words, part of sentences or whole sentences and does not examine the actual difficulty.

A traditional way of testing the effect of text simplification is by measuring differences in reading comprehension between an original text and a simplified text. To understand what type of simplification that can constitute better reading comprehension it is first necessary to explore what rules to base the simplification on. In this study the text simplification rules will be explored through words and sentences, not snippets or longer text segments. This will be carried out manually and with support of some automatic text simplification. Testing synonym replacement on difficult words and applying rules for semantic simplification has been done and proven effective (Aluisio, Specia, Gasperin, & Scarton, 2010) (Keskisärkkä, 2012) (Kandula, Curtis, & Zeng-Treitler, 2010). The idea is that this could provide useful insight into defining and testing existing rules of a text simplification software.

The text simplification software partially used throughout this project is based on a previous paper that uses a simplification called *StilLett* (Falkenjack, Rennes, Fahlborg, Johansson, & Jönsson, 2017). Based on StilLett some modifications were made before this study to improve some errors in output. This paper will not discuss the technical aspect any further, more information about the model can be found in the original paper.

Method

Background

Since there has been little research on how to evaluate Swedish text simplification software with user-based testing, this test has been developed with the purpose of exploring what such a test could look like. A goal with the test is to find out how it could be generalized to test a larger number of subjects. This requires the test to be adapted to individuals with ID. Findings from an ethnographic study on surveys aimed towards people with autism and intellectual disability presents seven important aspects to consider when designing a survey (Nicolaidis, o.a., 2020). Some of these could not be included in this design since they contradict the aim of the test. For example, not using "complex sentence structure, confusing grammar, or incomplete phrases" is inevitable when presenting two options during the test, one being the

original sentence, the other one being a simplified example. However, avoiding imprecise response options and confusing phrasing when presenting a question, can still be carried out. Also, decreasing anxiety levels by providing options of continuing without choosing an answer alternative if the question is perceived as too difficult.

The test is also inspired by insights and methods defined by the Swedish organization called *Begripsam* during their project on *Understandable Text* (begriplig text) (Johansson S. , 2019). During this project they developed an application called *Compair* that presents two alternatives against each other and instructs the subject to choose the most meaningful alternative. The thought behind comparing alternatives to each other is to create a joyful test experience and minimize workload. This project also discusses the importance of applying a qualitative method to understand the correct goals and using a developed prototype to aid user reflection.

Design

The study consisted of two iterations of a pilot test, the first iteration being an explorative test and the second being an additional test based on the insights gathered from the first test. Both tests were materially and procedure wise carried out in the same way and only differing in test content and participants. The test consisted of a survey with a total of 45 questions, 25 were lexical oriented questions, 20 were sentence oriented questions. Five of the lexical questions were towards and paraphrases of the words. When making the survey, an online software called PsyToolkit was used (Stoet, 2010, 2017).

Data

The data for the survey was constructed in a similar manner as a study on text simplification to support reading comprehension for children, where both operations such as splitting of sentences and lexical simplification were carried out (Belder & Moens, 2010). Lexical simplification can both refer to substitution of difficult words and paraphrasing. For this study, all questions were based on sentence data gathered from a Swedish geography book for students in classes seven through nine in elementary school. For the word part of the test, data was handpicked from the results of an automatized data collection, that identified less frequent words based on a corpus from the Swedish national language bank (*nationella språkbanken*). After a difficult word was identified, the sentence which the word was a part

of was also documented to provide context of meaning for each word. Each difficult word was checked for synonyms in a database called *Folkets synonymlexikon Synlex* that displays how much two synonyms relate to each other in meaning (Kann, edited: 2020). Folkets synonymlexikon is based on over 250000 answers, all graded with a value between 0-5, where close to zero is a low synonymity and close to five is a high synonymity. Not all words had synonyms in the lexicon and thus those words were ascribed synonyms from *synonymer.se*, a Swedish database of synonyms without information about synonymity. Synonyms from this website were selected manually with personal consideration. Each word was also automatically checked for potential taxonomical superordinate or subordinate words based on data from *spraakbanken.gu.se*. Superordinate words are often more general and abstract than the original word, while subordinate words are often more specific and explicit than the original word (Ungerer & Schmid, 2006). Lastly, all collected words and synonyms were noted with a usage frequency gathered from the corpus SUC 3.0 (*Stockholm-Umeå-korpus 3.0*), a Swedish corpus consisting of 1 66 593 tokens and 74 245 sentences (SUC 3.0, viewed: 2021). All of this was gathered in an Excel-file with token length and part of speech tag for each word (Appendix 1).

Collecting data for the paraphrasing questions were carried out in a similar manner as the word data during the process of identifying difficult words. As opposed to the word data, the paraphrase words were always a conjunction of two words. This made it easier to create a paraphrase by splitting the word and using a descriptive phrase to describe the original word. For example, the Swedish word *luftföroreningar*, which is translated to *air pollution*, is a conjunction of the words *luft* (air) and *föroreningar* (pollution). This word can be paraphrased as *föroreningar av luft*, which translates as *pollution of air*. Since there is no database for this, similar to those for synonyms, this had to be carried out manually with personal consideration. This was also collected and stored in a separate Excel-file (Appendix 1).

When collecting sentence data through StilLett to identify sentences and modify them based on a set of rules summarized in a paper based on the method of Rennes (2015) (Johansson V. , 2017)(Figure 1). Out of 20 modified sentences, six of them were based on the P2A (passive-to-active) rule, five were based on the SWO (straight word order). Since the StilLett is not finished, there were some modifications done to the sentences, mostly grammatical and structural. The remaining nine sentences were modified by rules of splitting a sentence into

two more sentences. These nine rules consisted of three SPLIT-k (splitting for subordinating and coordinating conjunctions), three SPLIT-r (splitting for relative clauses), and three SPLIT-a (splitting by an apposition). For these three rules the software was triggered on the part of a sentence where the split should be done but unable to perform the correct modification result. Therefore, the sentence split modifications were carried out manually based on the software's triggers. These were collected and stored in a separate Excel-file (Appendix 1).

Rule	Definition
Proximization (Prox.)	The rule aims to change the text to make it psychologically closer to the reader. This can be done by directly addressing the reader. This was done by changing the indefinite pronoun <i>man</i> (eng: <i>one</i>) to <i>du</i> (eng: <i>you</i>). Also, the correct form of the object corresponding to the pronoun was set, if needed.
Passive-to-active (P2A)	The rule aims to rewrite sentences of passive form to active form. The rule is triggered by a verb with the feature <i>SFO</i> , indicating a verb in passive tense.
Quotation inversion (QI)	The rule aims to change the place of a quotation and the person expressing it. The rule is triggered by sentences of quotation-like form. That is, a quotation followed by a comma, a verb, and a pronoun or a noun. A quotation either starts with a dash or has a quotation mark before and after the quote.
Straight word order (SWO)	The rule aims to rearrange the words in a sentence to achieve straight word order. That is, first a subject, then a verb, and then an object.
SPLIT-k	Sentence splits aims to divide long and/or complex sentences into new, simpler sentences. SPLIT-k performs splitting for subordinating and coordinating conjunctions. The rule was triggered by a comma followed by a word with POS-tag <i>SN</i> or <i>KN</i> .
SPLIT-r	A second split rule, which performs splitting for relative clauses. The rule was triggered by a relative pronoun (POS-tag <i>HP</i>) in a nominal phrase.
SPLIT-a	A third split rule, which performs splitting for appositions. The rule was triggered by an apposition (dependency label <i>AN</i>) within commas.

Figure 1. Shows the different rules of sentence simplification.

The first test

The questionnaire

In order to test what types of questions that would best fit for a finished test in the future, this test consisted of different question designs. Different structures were discussed when creating the test in effort to both establish understanding and advocate self-reflection in the subject.

The word part of the test was divided into four different question designs with five questions for each design. All questions required operation of either replacing an original word or keeping it. Each question consisted of a short text snippet to lower the perceived difficulty (Leroy, Endicott, Kauchak, Mouradi, & Just, 2013). The first five questions presented a sentence with the selected difficult word highlighted in blue with three alternatives below (Appendix 2). Two of the alternatives were possible synonyms and the third alternative was keeping the original word. Participants were verbally asked to choose the least difficult word alternative. Upon selecting an alternative, the following question displayed a prompt asking the subject to explain the word out loud.

The second group of five questions had the same design as the previously mentioned layout, but with the question *is this word difficult?* at the top of the page and with the alternatives *yes* and *no*, instead of possible synonyms (Appendix 2). Answering *yes* brought the subject to a follow up page with a prompt asking if there is an easier alternative to the original word, also providing three alternatives beneath. Two of the alternatives were synonyms and one was giving the option of selecting none of the alternatives. If one of the two synonyms were selected a follow up prompt encouraged the subject to explain the word out loud. The same follow up prompt was presented, but asking for an explanation of the original word, if the subject answered *no* on the first page.

Next part of five questions asked the subject to fill in a word gap of a sentence (*which word do you think fits in the gap?*). A sentence was presented with the selected difficult word replaced by an empty lined gap, followed by answer alternatives (Appendix 2). One of the alternatives being the actual word of the sentence, followed by alternatives being synonyms. Upon choosing one of the alternatives, the subject was asked to explain the word out loud.

The last answer alternative was selecting none of the alternatives, stating that no word fits in the gap. This directs the subject directly to the next question.

For the last part of five questions, the layout consisted of a question requesting the user to choose the easiest word, but this time, with no example sentence. Instead, the answering alternatives were each a sentence with the manipulated word highlighted in blue (Appendix 2). One of these alternatives was the original sentence and the others the same sentence but with a synonym to the difficult word highlighted in blue instead. As with the other parts, when choosing an alternative, the subject was directed to a prompt asking them to explain the highlighted word out loud. The subject was also presented with an answer alternative to continue without choosing an alternative, thus directly jumping to the next question.

All the questions for the sentence part had the same design. They consisted of a question asking the subject to choose the easiest sentence and was followed by three answering alternatives, one being the manipulated sentence, one being the original sentence and the last continuing without choosing an alternative (Appendix 2). Which one of the two sentences that was presented as the first/upper alternative was randomized for each question. This randomization was implemented to make sure that the subject did not just choose the same alternative by order for every questions. When choosing the alternative of continuing without taking a stance, this would bring them directly to the next sentence question. However, if a subject selected one of the sentence alternatives, they were then directed to a page where they could rate the simplicity of the chosen sentence on a scale from one (being a little easy) to seven (being very easy). After rating the sentence there is a last page that invites the subject to reflect on the sentence and doing this out loud. This layout goes on throughout all 20 sentence questions.

The last part of the test consists of five questions about paragraphing. Although also being lexical oriented as the synonym part, the paragraph questions were separated from the word questions to look for potential different effects. Much like the sentence questions, the subject was asked to choose the simplest of two alternatives and was provided with an alternative of continuing without choosing. Each question had an alternative in form of a sentence with the potential difficult word highlighted in bold text and a second alternative being the same sentence but with the difficult word paraphrased, also highlighted in bold (Appendix 2). If the subject chooses the sentence alternative with a word or the alternative with a paraphrase,

the next screen asks the subject to explain out loud the word or the paraphrase, depending on what has been chosen. Choosing the alternative of continuing without selecting either of the alternatives directs the subject to the next question.

Participants

Two participants agreed to do take part of the first test. Both were over 18 years old males and were studying their fourth year of the special class in nature and social sciences program at a high school in a large Swedish city. The first participant (P1) reported sufficient reading skills and the second participant (P2) reported having difficulty with reading and being a slow reader, thus requiring someone to read out loud.

Both participants were informed about management of the test data and their right to end the test or withdraw their participation at any given time. Before starting the test both participants reported consent of receiving information about data collection and participation, orally and by checking in a box in the test. The participants parents were informed about their participation. One teacher was also present throughout the whole test and helped P2 with reading.

Procedure

The study was carried out in a secluded quiet room at the school with only a computer present, requiring nothing else than internet connection. Each participant got to state their name, gender, age, year of class and level of reading skill. After having received information about data collection and about the test, the subject was asked to begin the test. Throughout the test, notes on the reflective questions were gathered with a paper and pen. After each of the three mayor test parts the subject was asked a couple of questions in a semi structural manner. They were asked questions about attitude towards the specific part of the test, if any aspect of the test was more difficult or if any aspect was more appealing. These questions were not scripted since understanding and social skills varies between different individuals with ID. At the end of the test, each subject was asked to reflect on the test as a whole and if any parts where more difficult than others. They were also asked to share any other reflections about the test. When everything was done the participant was thanked and returned to class.

Data collection

All data was automatically stored with PsyToolkit and later put and processed in a Excel-file. The notes collected on paper were reprocessed in a computer document and restructured for easier reading.

Results

Quantitative data

The quantitative results from the word part of the test show that both participants answered all the questions. P1 choose the first alternative 30% of the time, the second alternative 45% of the time, the third alternative 25% of the time and the fourth alternative 0% of the time. P2 choose the first alternative 30% of the time, the second alternative 35% of the time, the third alternative 30% of the time and the fourth alternative 5% of the time. This indicates that there was no bias towards choosing a specific alternative for every question (except the fourth alternative which was the option of continuing without choosing an alternative).

On average, P1 choose a higher number of original words and choose generally less frequent words (Table 1). Both participants choose somewhat equally long words and only choose superordinate or subordinate alternatives 10% of the time.

Table 1. Data showing results from word-part of the test. Average word frequency occurrence (from SUC 3.0) show that P2 generally chose twice as common words than P1. P1 also chose the original word more often than P2. Synonymity is scored on a scale from 1-5, where most words are labeled with three or higher, which makes 3 and 3,7 not that different.

	P1	P2
average word frequency	5,02	10,18
average synonymity	3	3,7
%words original	75%	52,63%
superordinate	5%	10%
subordinate	5%	0%
average word lenght	8,75	9,05

50% of the questions (ten questions) where answered with the same alternative for both participants. For word part 2, with possibility of replacing current word in sentence, P1 kept 100% (five words) of the words and P2 kept 60% (three words) of the words. Only one time was the alternative of continuing without choosing a word carried out, which was during P2s word part 3.

The quantitative results from the sentence part of the test show that both participants answered all the questions. P1 choose the modified sentence 65% of the time, the original sentence 35% of the time and had a mean value of 4,95 on attitude towards difficulty of sentence (seven being very easy and one being very difficult) (Figure 2). P2 choose the modified sentence 20% of the time, the original 80% of the time and had a mean value of 5,45 on attitude towards difficulty of sentence. P1 choose the first order alternative 40% of the time and the second order alternative 60% of the time. P2 choose the first order alternative 50% of the time and the second order alternative 50% of the time. This shows no bias of choosing one order alternative over the other, thus making the difference in answers between participants depend on preference of modified or original sentences. None of the participants choose the alternative of continuing without choosing a sentence.

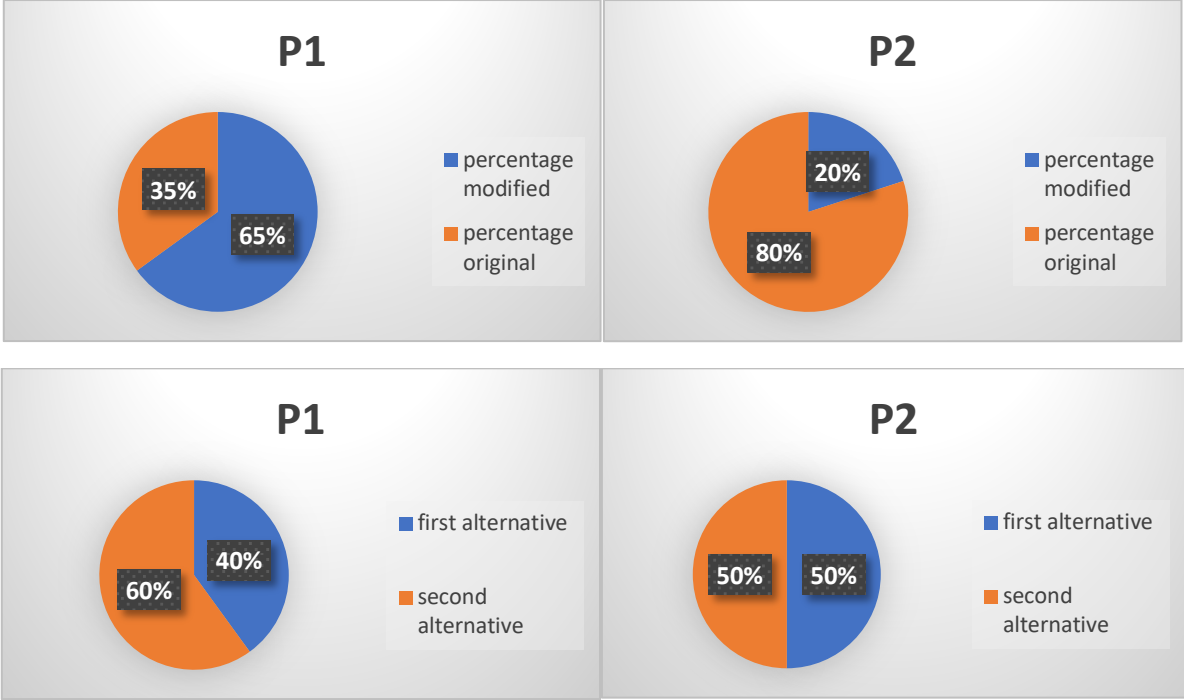


Figure 2. Data showing results from sentence-question part of test. The two upper graphs show percentage of chosen sentences that were modified and original for both P1 (to the left) and P2 (to the right). The lower graphs show percentage of chosen alternatives order for both P1 (to the left) and P2 (to the right). Order does not seem to be relevant to if the participant chooses a modified or an original alternative more.

The quantitative results from the paraphrase part of the test show that both participants answered all the questions. P1 choose the modified alternative 20% of the time and the original alternative 80% of the time (Table 2). P2 choose the modified alternative 60% of the time and the original alternative 40% of the time. P1 answered 40% of the questions by choosing the first alternative, 60% by choosing the second, and P2 answered 100% of the questions by choosing the second alternative. Since this part only contained five question it's

difficult to speculate if there is a bias towards choosing the second alternative. None of the participants choose the alternative of continuing without answering.

Table 2. Data showing results from paragraph-part of test. Upper table show chosen type and lower table shows type of input.

parafras		
Fråga	T1	T2
Q1	original	modified
Q2	modified	modified
Q3	original	original
Q4	original	original
Q5	original	modified
Fråga	T1	T2
Q1	1	2
Q2	2	2
Q3	2	2
Q4	2	2
Q5	1	2

Qualitative data

During the test P1 expressed good confidence, did not ask for any help and did not express any overall opinion about the test. When asked to compare the different parts of the test, P1 had nothing to add besides that the sentence part required a bit more concentration.

For the second participant P2 the test was more difficult. P2 was able to understand and answer all the questions but required help with reading. After the test P2 did not express any complaints about the test nor perceiving any difficulty. When asked to compare the different parts of the test, P2 reported that the sentence part required more concentration and that some questions contained too much text.

Discussion

Although the results are overall positive there are a couple of things worth discussing about the test. To begin with, the screen that asked the subject to reflect on the question, for example *“is the highlighted word difficult? (Explain out loud)”*, was very ineffective. Most of the time it did not generate any reflection and those occasions that it did were often followed by a misunderstanding of the question. When asked to explain the chosen word for a question,

P1 tended to use one of the synonym alternatives from the question or paraphrasing the word. P2 seemed to have a difficult time answering the reflective questions, taking an extensive amount of time to answer, or just continuing without answering. After a couple of reflective questions, both participants were asked to skip those question.

The results from the interview questions during and after the test show that none of the participants expressed difficulties with the tests. However, this may not be the case at all since completing the test itself is easy. The real question is whether the test was carried out in a way that that answers the research questions. For instance, P1 seemed shy and he may have focused more on performance which may have limited his ability to give feedback on the test. The reason for this inference is based on how P1 answered word part 2. During the second word part, P1 kept the original word for every question which was not the case during the other parts. This may be explained by the fact that word part 2 requires the subject to take a stance to whether the presented word is difficult or not before having alternatives presented and thus not being able to associate the word. Previous findings have shown that binary answering alternatives in questionnaires are not optimal since they often lead to compliant answers from subjects with ID (Stock, Davies, & Wehmeyer, 2004). But in this case, it seems to be the opposite, since both P1 and P2 reported a question being difficult at least one time each. P1 may have been opposed of explicitly admitting to a word being difficult. This could also be the reason why P1 never choose the alternative of continuing without answering a question, because that would be the same as admitting to not understanding the alternatives. But at the same time, it could just be that P1 experienced the test being easy. This is difficult to know since P1 was not very talkative during interviewing and reflection. It would have been interesting to have an alternative word in the presented sentence for a couple of the questions in word part 2, to see if the participant still would choose to keep each word. The most probable explanation could be that since there are no alternatives presented during the first screen it could be problematic to answer whether the word is difficult without being able to compare it with easier or more difficult alternatives. Indeed, the second word part could involve a too high workload, having a too complicated structure where the user first must compare a certain word to their own vocabulary and only when answering yes be provided with potential better alternatives.

A similar concern can be found in the answers that P2 provided during the sentence part of the test. During the screen where the subject was asked to rate the difficulty of a sentence on a scale from one (very difficult) to seven (very easy), P2 never answered below four. As with P1 during the word part, this could be a result of P2 experiencing the sentences as being easy. But it could also be the result of resisting to admit that a question is difficult. This became more evident when P2 was interviewed about the difficulty and answered that the test was easy because he managed to answer all the questions. This shows that P2's main goal was completing the test without failure, even though the test is impossible to fail and has no right or wrong answers. When P2 gave each sentence a score of four or higher, it could have been an answer to *if the question were difficult to answer* rather than *if the sentence itself was difficult to understand*. Having a rating is not optimal without asking any direct related questions to why the subject gives a certain score.

When P1 and P2 was explicitly asked which of the answering formats they perceived as easiest they first did not answer. But when presented with each of the word parts structures, both were more positive to the "fill in the gap" structure in word part 3. At the beginning, P2 seemed to have some problems with answering the questions in the first word part. No reflective answer was provided so it is difficult to answer what about the question structure of the first word part that was difficult. One explanation could be that it requires opposing the established status quo since the sentence is already complete and the subject is asked to modify it. Studies of compliance level for prefilled answers in questionnaires show different findings. But an effect of higher compliance when there is a correct prefilling has been found in a few studies on tax forms (MartinFochmann, Müller, & Overesch, 2021). Since none of the word alternatives are wrong and if the participant understands the presented word in the sentence, there could be lower incentive for the subject to choose an alternative word. Another explanation could be that answering questions with a word structure as in the first word part, could require too many cognitive operational steps. This could be represented as the subject first needs to read the sentence, focus on the highlighted word, read the alternatives, mentally put an alternative in the sentence, read the sentence again with the alternative word and compare it to the original sentence. Although some of these steps seem simple or redundant, they may be crucial for an individual with ID to correctly compare words. This is suggested when looking at the screen recording of P1 and following his mouse cursor

movement. It is observable that P1 moves the mouse cursor to help follow along during reading and doing so in a similar way as the operations are described to be carried out above. Thus, word part 1 may have a too difficult question structure.

Lastly, further on the test should be carried out on subjects that do not require any reading help. This became evident during the sentence part with P2 when different splitting of sentences was not conveyed through listening to the alternatives. Worth mentioning is that although P2 had difficulties with reading, he was able to perform comparisons and necessary decoding of the alternatives once he had the text read to him. In other words, P2 was able to connect the audial information to correct textual information and then process the textual information without further help. This could of course affect the overall performance of the test in other ways, for example minimizing possible misunderstandings from failure of decoding complicated or unfamiliar words. But the most prominent case was with comparing alternatives of sentence split.

The second test

The questionnaire

For the second test a couple of adjustments on the same questionnaire were made based on the insights from the first test. The beginning of the test consisted of a few questions about general information such as gender, age, school, level of studies, participant alias and confirmation of received information about consent (Appendix 3). This was done so that future test can more easily gather all general in the same place.

Some parts of the first test were removed to isolate and better understand the positive perceived aspects of the test. As presented in the discussion of the first test both structures of word parts 1 and 2 were speculated to be more difficult than word parts 3 and 4. For this reason, both the structure with replacing a highlighted word and beginning with answering if a word is difficult or not before getting alternatives, was removed. Therefore, the questionnaire consisted of two word parts with ten questions each instead of four word parts with five questions each.

Since there were some misunderstandings and issues with the reflective questions, these had to be change as well. In the previous test, subjects were asked to explain the chosen word in the word part, which did not initiate any meaningful reflection. Instead, this test asked the

participant to explain why they choose a specific word (Appendix 3). This in effort of creating reflection that revolves around comparing presented alternatives and gaining insight to why a subject feels that a certain word better fits the sentence context. The same adjustment was made for the paragraph part, asking participants to explain the selection rather than explaining the word or phrase. For the sentence part, the reflection question was changed from *what did you think about the sentence* to *why did you choose said sentence*. To reflect about feelings towards a sentence may be more abstract and complicated to answer than to motivate the thought behind choosing a sentence.

The rating scale was also changed since the ratings of sentences showed skewed results, where sentences only received a four point or higher. Indeed, Likert-format scales tend to be misunderstood by individuals with ID and are therefore not very effective in these types of questionnaires (Stock, Davies, & Wehmeyer, 2004). Instead, the follow up screen presented a question asking if the sentence was difficult or not (Appendix 3). Rating a sentence on a scale from one to seven is more complex and less effective than providing a binary answer, even though, as discussed above, binary questions can result in acquiescent answers.

More randomizations were also implemented. The order for all answering alternatives were randomized, for every part of the test, so that the original expression did not necessarily have to be first in the order. There was also a randomization of order between questions implemented. This was done so that the subject would not be able to expect upcoming questions and thus answer having to change processes between questions. A second reason was that the future questionnaire will most likely have to contain a randomization of questions to prevent confounding variables. Doing this in the pilot study can provide insight to if this establishes problems or not.

Participants

One participant agreed to take part of the second test. The participant was a 17-year-old female and a student of special class in nature and social science program at a high school in a large Swedish city. This participant (P3) reported good reading skills requiring no extra help. Before the test, the participant was informed about management of the test data and their right to end the test or withdraw their participation at any given time. A consent form was signed, and the participants parents were informed about the test and their right to withdraw the participant from the test. The participant also checked the box in the beginning of the test,

declaring to understand and have received information about data collection and participation.

Procedure

The same procedure was carried out in this second part of the study.

Data collection

All data was automatically stored with PsyToolkit and later put and processed in a Excel-file. The notes collected on paper were reprocessed in a computer document and restructured for easier reading.

Result

Quantitative data

The quantitative results from the word part of the test show that the participants answered all the questions. P3 choose the first alternative 55% of the time, the second alternative 35% of the time and the third alternative 10% of the time (Figure 3). This shows a slight bias towards choosing the first alternative. The answers alternatives consisted of 45% words from the original data, 15% of words from *synonymer.se*, 20% from *folkets lexikon*, 15% superordinate words, and 5% subordinate words.

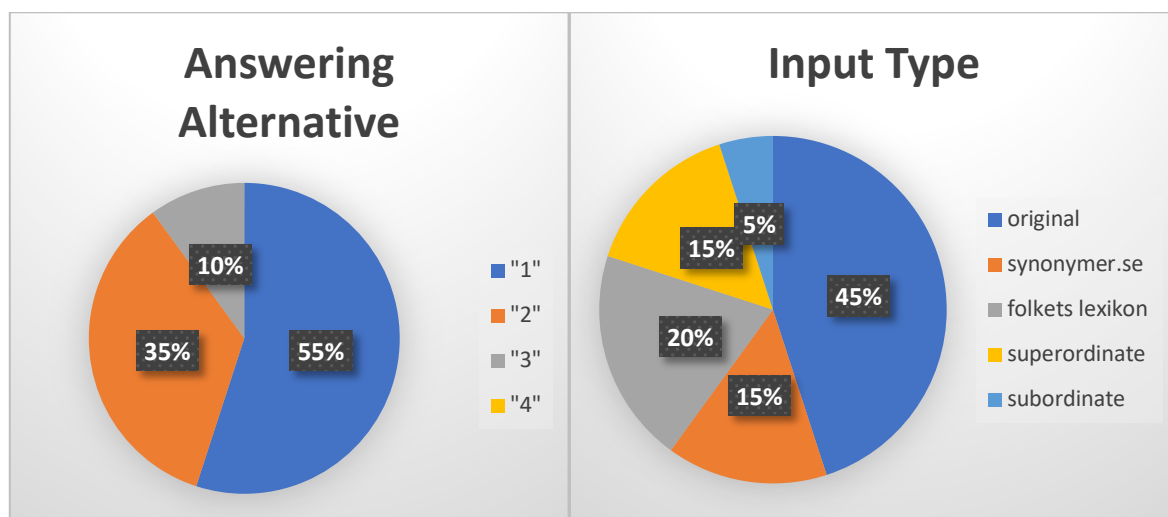


Figure 3. Left graph shows the distribution of input in percent. Right graph shows distribution for type of chosen word in percent.

From the data on occurrence frequency in *SUC 3.0*, a relative frequency was calculated. The relative frequency had a mean value of 0,22 where values closer to 1 are considered words more frequent relative to their word alternatives (Figure 4). This contradicts the hypothesis of

the *lexical quality hypothesis*, where more frequent words should be preferred (Perfetti & Hart, 2002).

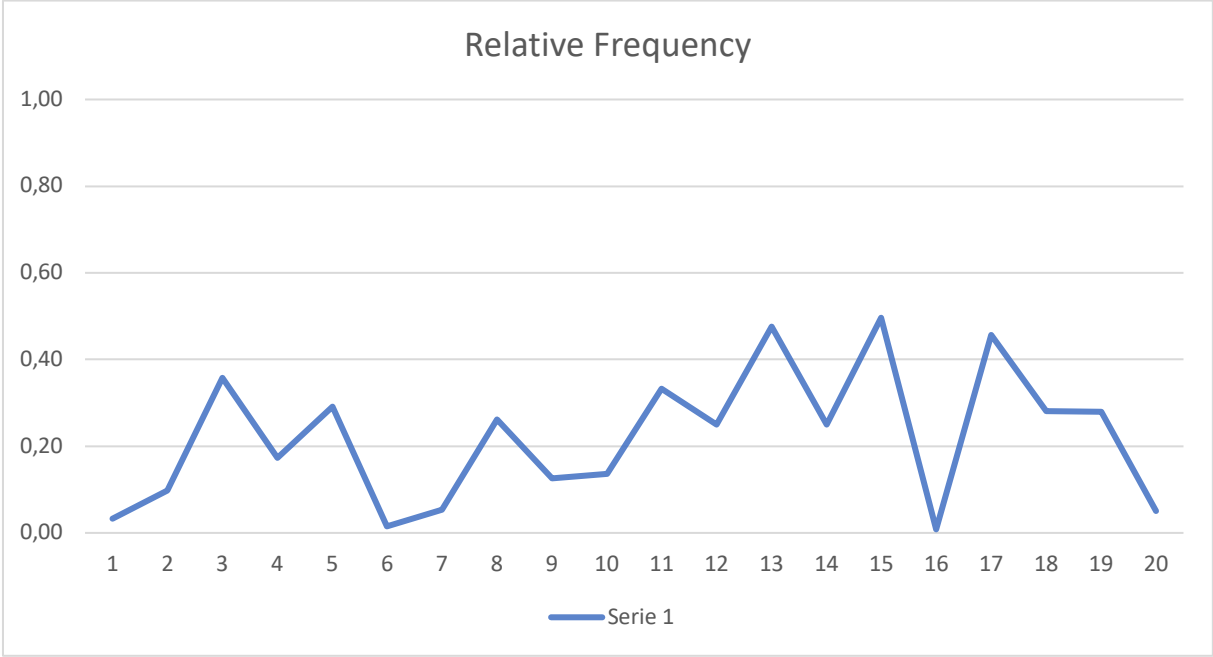


Figure 4. A graph displaying relative frequency of a word's occurrence, where closer to one is a relative more frequent used word and closer to zero is a relative less frequent used word.

A graph over the relative word length was also plotted with a mean of 0,37, where values closer to one are considered longer words relative to their word alternatives (Figure 5). These results show that P3 is more likely to choose a shorter word than a longer word.

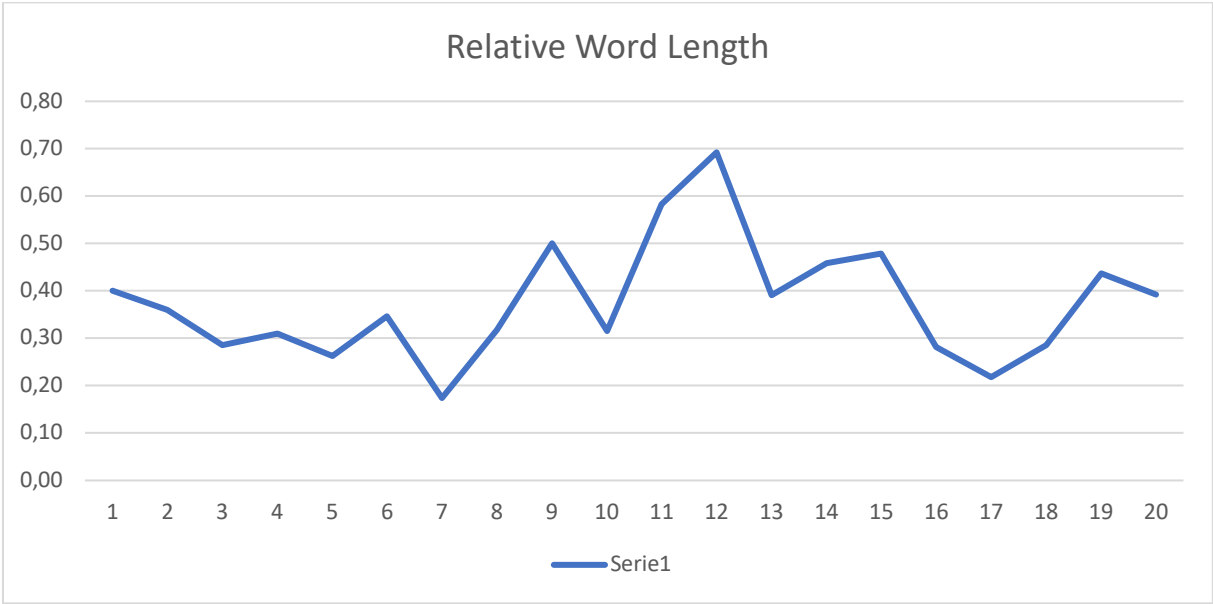


Figure 5 graph displaying relative word length of a word, where closer to one is a relative longer word and closer to zero is a relative shorter word.

The qualitative results from the sentence part of the test show that the participant answered all the questions. P3 choose the modified sentence 40% of the time, the original sentence 35% of the time and continuing without choosing an alternative 30% of the time (Figure 6). P3 choose the first order alternative 35% of the time, the second order alternative 35% of the time and the third order alternative 30% of the time.

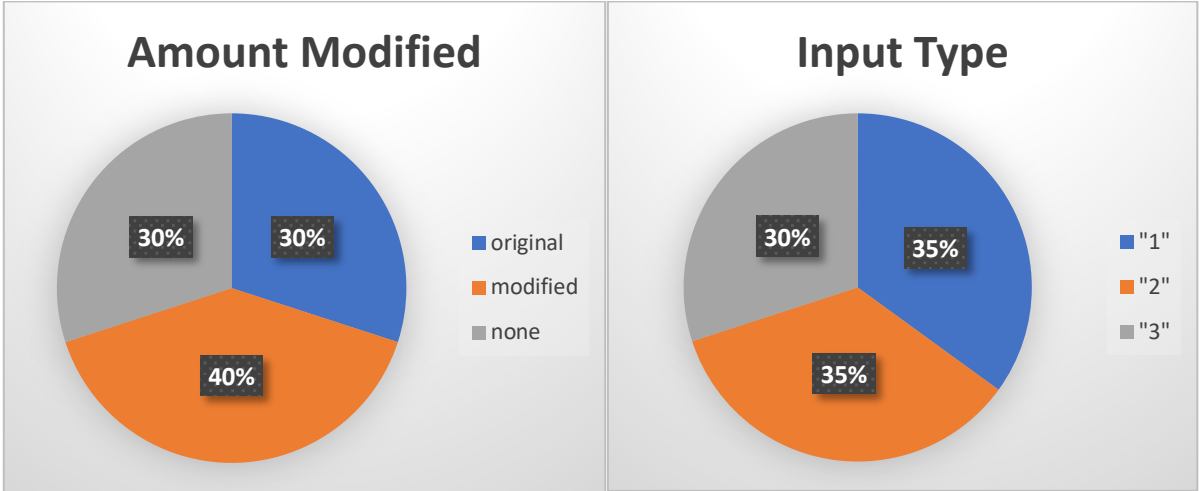


Figure 6. Left graph shows distribution over type of chosen sentence in percent. Right graph shows distribution over input type in percent.

The answers from the binary questions about difficulty shows that 83% of the original sentences were considered easy and that 75% of the modified sentences were considered easy (Figure 7).

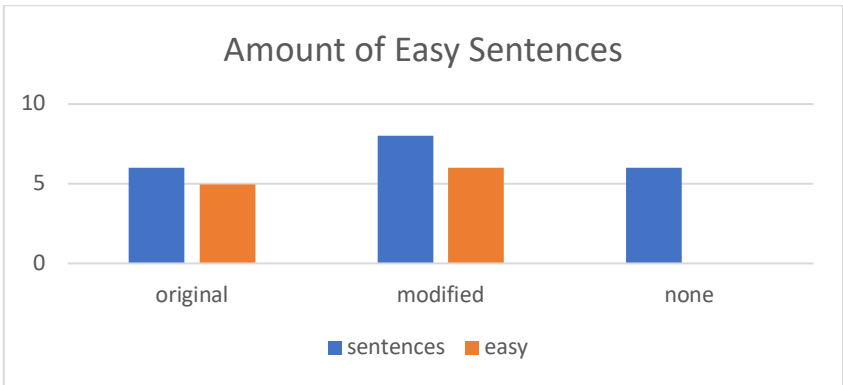


Figure 7. This graph displays how many of the chosen type sentence were considered easy in comparison of the total amount.

The qualitative results from the paraphrase part of the test show that the participant answered all the questions. P3 choose the original alternative 60% of the time and the modified alternative 40% of the time (Figure 8). P3 answered 60% of the questions by choosing

the first alternative and 40% by choosing the second. None of the answers were choosing the alternative of continuing without answering.

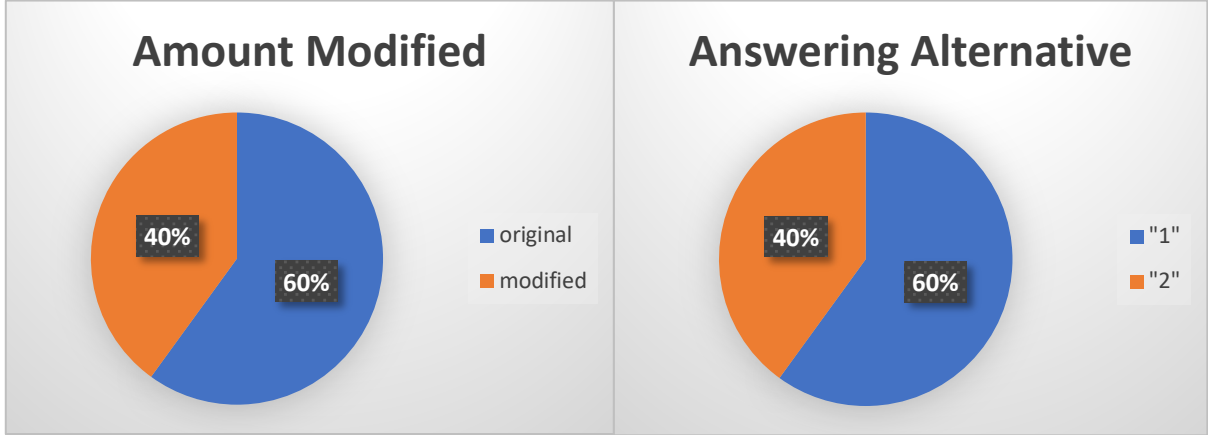


Figure 8. Left graph shows distribution over amount of chosen modified or original phrases in percent. Right graph shows distribution of input in percent.

Qualitative data

During the test P3 expressed good confidence, did not ask for any help and did not express any overall opinion about the test. P3 reported a mixed perceived difficulty of the test and that the word-oriented questions were easiest. The test was not perceived as too long, and the instructions were easy to follow according to P3. No further comments were added when P3 was asked if she had any own questions or thoughts about the test.

Based on observations, the participant seemed to become bothered by too many reflective questions after a certain amount of time. This was most evident when P3 began progressing to the next screens without answering the present reflective questions. However, P3 did answer more reflective questions than both P1 and P2.

Some of the reflection questions generated interesting results. For instance, asked why P3 choose the word *permanently* to fill the gap in the sentence “*There are no people ____ living in Antarctica*, she answered “*well, because nobody wants to live there*”. Which indicates some sort of misunderstanding.

The reflection from P3 during the sentence-oriented questions of the test were not as easy interpreted as for the word questions. For instance, P3 had difficulties motivating she choose one sentence over another. This was partly due to the self-answering expression of the reflection question presented after the selection of a sentence. Since the sentence question requests the subject to choose the easiest alternative, it becomes obvious that the answer to

the reflective question “*why did you choose this sentence?*” is “*because it is the easiest question*”. This is exactly how P3 answered the questions.

Discussion

As the case with the first test, it is difficult to know if the results are as positive as they appear. Although the overall procedure of the second test was more refined and worked better with the participant, there are some aspects worth discussing. For instance, the increasing aggravation with the reflective questions parallel to test progression. P3 was more inclined to answer a larger number of reflective questions than both P1 and P2, which may have been a result of rephrasing the reflection questions after the first test. This positive effect was limited and P3 became notable distracted and irritated after a while. A potential solution is decreasing the number of reflective questions and spreading them out, so that they do not appear sequentially.

It is interesting to note that there seem to be no clear correlation between word frequency and chosen word or word length and chosen word. There is too little participant data and too few alternatives to make a statistical statement of the result and therefore can the result only be speculated. The graphs would probably look differently if the total frequency and word length were the same for all questions. Analyzing the means of the two graphs show that P3 generally preferred words with lower occurrence frequency and shorter word length. Choosing lower frequent words instead of more common words could depend on personal characteristic factors, such as vocabulary size and domain knowledge, which influences the perceived difficulty of a word or text (Leroy, Endicott, Kauchak, Mouradi, & Just, 2013).

The selection of less frequent words could also be explained by the decoding of the word in each sentence context. According to the Reading Systems Framework, cognitive processes between semantical and lexical knowledges results in context specific word meaning during decoding (Perfetti & Stafura, 2014). This can be backed up by the misunderstanding, presented in the results, that occurred during P3s reflection on the question about people living on Antarctica, which was answered literally. Not explaining why she choose the specific word, could be interpreted as she being fully occupied with the meaning of the sentence and choosing the word best fitting the context. Even though P3 may have rated *constantly* or *forever* as easier alternatives than *permanently*, she could have chosen the more difficult

alternative because the context of the sentence prescribed meaning or understanding to the word. This is interesting since it highlights a problem with the *lexical quality hypothesis* which advocates substitution of less frequent words with more frequent words to aid reading comprehension (Perfetti & Hart, 2002). Although this still may be the case, it is necessarily not something that the reader itself would prefer or choose if given the alternative.

The problem with ill-defined reflection questions during the sentence part of the test was temporarily resolved by asking more thorough questions, asking P3 why the chosen sentence was perceived easier than the other. This follow-up question did still not provide more nuanced answers. P3s two main motivations for choosing one sentence over another were *“both are difficult sentences, but this one was a little bit easier”* and *“I choose the one that is most obvious”*. It seems that selection of sentences is mainly carried out with intuition, which likely prevents any meaningful meta cognitive reflection.

It is still unclear if different sentences are compared in the expected way. Highlighting modified components and changes with both bold and blue seem to have worked well in the word and paragraphing questions for both tests, since they have been interpreted in the correct way. The problem is that there is no such visual guidance for the sentence part. This becomes more problematic since visual guidance is shown to improve understanding in people with ID (Davies, Stock, & Wehmeyer, 2003). This is something that could be explored further, by for example highlighting punctuation for split sentences or putting rephrased part of sentences in bold. Of course, there is a risk that not the entire sentences are being compared, but only the modified parts.

Conclusion and further work

This study has explored a potential test design for evaluating text difficulty on a word- and sentence-level, specifically for individuals with ID. Developing such a test could lead to further understanding of reading comprehension and potential automated aids. Some scientific background has been provided and has highlighted the lack of research on ID and reading comprehension. By applying a qualitative research method, this paper has provided knowledge in test design preferences and perceived text difficulty with ID adolescents as target group. Although the results and discussion does not conclude any fixed solutions or complete test design, they do present less effective design options and underline problematics

with meta cognitive reflection. The research question *how can we examine self-reported comprehension and textual preferences in adolescents with ID?* is answered to some degree. As presented in this paper, some meaningful scientific knowledge has been found, but more exploration is required. At this point it is not possible to conclude if a test carried out on a broad population would produce statistic significant results. Personal preference and knowledge differences between individuals with ID could be so immense that creating a general test simply is impossible.

Hopefully, this paper will be applicable for future test development and research in this area. Suggestions for future work is to proceed on this work to develop a test that can be carried out on a small population that will provide generalizable quantitative data that can be fed back into an automatized text simplification program. Also expanding exploration of reading comprehension tests to other atypical readers, such as individuals with dyslexia.

References

- Aluisio, S., Specia, L., Gasperin, C., & Scarton, C. (2010). *Readability Assessment for Text Simplification*.
- Arfe, B., Mason, L., & Fajardo, I. (2018). Simplifying informational text structure for struggling readers. *Reading and Writing*.
- Barker, R. M., Sevcik, R. A., Morris, R. D., & Ronski, M. (2013). A Model of Phonological Processing, Language, and Reading for Students with Mild Intellectual Disability. *American Journal on Intellectual and Development Disabilities*, 365-380.
- Belder, J. D., & Moens, M.-F. (2010). Text Simplification for Children. *Proceedings of the SIGIR workshop on accessible search systems*. New York: ACM.
- Broek, P. v., Helder, A., & Karlsson, J. (2014). A Cognitive View of Reading Comprehension: Implications for Reading Difficulties. *Learning Disabilities Research & Practice*, 10-16.
- Chen, P., Rochford, J., Kennedy, D. N., Djamasbi, S., Fay, P., & Scott, W. (2017). Automatic Text Simplification for People with Intellectual Disabilities. *Artificial Intelligence Science and Technology*, 725-731.
- Conners, F., J.A.Atwell, Rosenquist, C. J., & Sligh, A. C. (2001). Abilities underlying decoding differences in children with intellectual disability. *Journal of Intellectual Disability Research*, 292-299.
- Davies, D. K., Stock, S. E., & Wehmeyer, M. L. (2003). A Palmtop Computer-Based Intelligent Aid for Individuals With Intellectual Disabilities to Increase Independent Decision Making. *Research & Practice for Persons with Severe Disabilities*, 182-193.
- Definition of Intellectual Disability*. (retrivied: 2021, Mars 3). Retrieved from AAID: <https://www.aaid.org/intellectual-disability/definition>
- Dudley, A. M. (2005). Rethinking Reading Fluency for Struggling Adolescent Readers. *Beyond Behavior*, 16-22.
- Fajardo, I., Avila, V., Ferrer, A., Tavares, G., Gomez, M., & Hernandez, A. (2014). Easy-to-read Texts for Students with Intellectual Disability: Linguistic Factors Affecting Comprehension. *Journal of Applied Research in Intellectual Disabilities*, 212-225.
- Fajardo, I., Tavares, G., Ávila, V., & Ferrer, A. (2013). Towards text simplification for poor readers with intellectual disability: When do connectives enhance text cohesion? *Research in Developmental Disabilities*, 1267-1279.
- Falkenjack, J., Rennes, E., Fahlborg, D., Johansson, V., & Jönsson, A. (2017). *Services for Text Simplification and Analysis*. Linköping, Sweden: Linköping University and RISE SICS East AB.
- Hollomotz, A. (2018). Successful interviews with people with intellectua disability. *Qualitative Research*, 153-170.
- Hoover, W. A., & Gough, P. B. (1990). The Simple View of Reading. *Reading and Writing*, 127-160.
- Johansson, S. (2019). *Begripsams metoder och metoderna vi använt i projektet Begriplig Text*. Begripsam helt enkelt.

- Johansson, V. (2017). *Depending on VR – Rule-based Text Simplification Based on Dependency*. Linköping University: (Bachelor's thesis).
- Kandula, S., Curtis, D., & Zeng-Treitler, Q. (2010). A Semantic and Syntactic Text Simplification Tool for Health Content. *AMIA Annual Symposium Proceedings Archive*, 366-370.
- Kann, V. (edited: 2020, 2 21). *Folkets synonymlexikon Synlex*. Retrieved from <http://folketslexikon.csc.kth.se/synlex.html>
- Keskisärkkä, R. (2012). *Automatic Text Simplification via Synonym Replacement (Master Thesis)*. Linköping: Linköping university.
- Knox, M., Mok, M., & Parmenter, T. R. (2000). Working with the Experts: Collaborative research with people with an intellectual disability. *Disability & Society*, 49-61.
- Koizumi, M., Saito, Y., & Kojima, M. (2019). Syntactic development in children with intellectual disabilities – using structured assessment of syntax. *Journal of Intellectual Disability Research*, 1428-1440.
- Leroy, G., & Kauchak, D. (2014). The effect of word familiarity on actual and perceived text difficulty. *Journal of the American Medical Informatics Association*, 169-172.
- Leroy, G., Endicott, J. E., Kauchak, D., Mouradi, O., & Just, M. (2013). User Evaluation of the Effects of a Text Simplification Algorithm Using Term Familiarity on Perception, Understanding, Learning, and Information Retention. *Journal of medical internet research*.
- Llewellyn, G. (1995). Qualitative research with people with intellectual disability. *Occupational Therapy International*, 108-127.
- Lorch, R. F., & Broek, P. v. (1997). Understanding Reading Comprehension: Current and Future Contributions of Cognitive Science. *Contemporary Educational Psychology*, 213-246.
- MartinFochmann, Müller, N., & Overesch, M. (2021). Less cheating? The effects of prefilled forms on compliance behavior. *Journal of Economy Psychology*, 1-20.
- Nicolaidis, C., Raymaker, D. M., McDonald, K. E., Lund, E. M., Leotti, S., Kapp, S. K., . . . Zhen, K. Y. (2020). Creating Accessible Survey Instruments for Use with Autistic Adults and People with Intellectual Disability: Lessons Learned and Recommendations. *Autism in Adulthood*, 61-76.
- Paris, S. G. (2005). Reinterpreting the development of reading skills. *Reading Research Quarterly Vol. 40, No. 2*, 184-204.
- Pennequin, V., Igier, S., Pivry, S., & Gaschet, N. (2021). Metacognition and emotion in adolescents with intellectual disability: Links with categorization performance. *Current Psychology* .
- Perfetti, C. A., & Hart, L. (2002). The Lexical Quality Hypothesis. *Precursors of functional literacy*, 67-86.
- Perfetti, C., & Stafura, J. (2014). Word Knowledge in a Theory of Reading Comprehension. *Scientific Studies of Reading*.
- Rello, L., Baeza-Yates, R., Dempere-Marco, L., & Saggion, H. (2013). Frequent Words Improve Readability and Short Words Improve Understandability for People with Dyslexia. *Human-Computer Interaction – INTERACT 2013. INTERACT 2013. Lecture Notes in Computer Science, vol 8120* (pp. 203-219). Berlin, Heidelberg: Springer.

- Rennes, E. (2015). *Improved Automatic Text Simplification by Manual Training*. Linköping University : (Master's thesis).
- Shardlow, M. (2014). A Survey of Automated Text Simplification. *International Journal of Advanced Computer Science and Applications*, 58-70.
- Stoet, G. (2010). PsyToolkit - A software package for programming psychological experiments using Linux. *Behavior Research Methods*, 42(4), 1096-1104.
- Stoet, G. (2017). PsyToolkit: A novel web-based method for running online questionnaires and reaction-time experiments. *Teaching of Psychology*, 44(1), 24-31.
- Stock, S. E., Davies, D. K., & Wehmeyer, M. L. (2004). Internet-Based Multimedia Tests and Surveys for Individuals with Intellectual Disabilities. *Journal of Special Education Technology*, 43-47.
- SUC 3.0. (viewed: 2021, 5 14). Retrieved from Språkbanken text: <https://spraakbanken.gu.se/resurser/suc3>
- Ungerer, F., & Schmid, H.-J. (2006). *An Introduction to Cognitive Linguistics*. London: Routledge.
- Wingerden, E. v., Segers, E., Balkom, H. v., & Verhoeven, L. (2014). Cognitive and linguistic predictors of reading comprehension in children with intellectual disabilities. *Research in Developmental Disabilities*, 3139-3147.
- Wingerden, E. v., Segers, E., Balkom, H. v., & Verhoeven, L. (2016). Foundations of reading comprehension in children with intellectual disabilities. *Research in Developmental Disabilities*;
- Wingerden, E. v., Segers, E., Balkom, H. v., & Verhoeven, L. (2016). Foundations of reading comprehension in children with intellectual disabilities. *Research in Developmental Disabilities*.

Appendix

Appendix 1

ordklass	frekvens	ordlängd	synonym	frekvens	ordlängd	synonymitet	synonym	frekvens	ordlängd	synonymitet	överordnat/underordnat ord	frekvens	ordlängd	synonymitet
verb	1,71	12	föreställa	15,43	10		beteckna	9,42	8					
particip	14,57	9	speciella	24,86	9	4	bestämd	35,14	7					
adjektiv	1,71	5	ojämn	5,14	5		grov	17,14	4					
substantiv	8,57	12	tillstånd	16,27	9	3,3	egenskap	22,28	8	3				
verb	0,86	8	fälls	5	6	4,6	hugga	3,42	5	3,5				
adverb	2,57	9	ständigt	69,43	8	4,3	bestående	9,42	9	4,8				
substantiv	0,86	11	uttryck	48,86	7	4,3	term	6	5	3,2				
substantiv	2,57	11	land	115,72	4	3					området	129,44	7	4,2
substantiv	2,57	10									millennium	0,86	10	
verb	6	8	flödar	2,57	6	4	välla	0,86	5	3,2				
substantiv	1,71	7	hed		3						slätt	0,86	5	
substantiv	1,71	9									regn	1,71	4	
substantiv	0,86	9	skada	18,86	5	4	katastrof	17,14	9					
substantiv	0,86	11									bottenslammet	0,86	13	
substantiv	0,86	12									anledningar	119,15	11	
substantiv	1,54	9	ställning	91,72	9	4,3					klippformation	0,86	14	
substantiv	0,86	8	bottensats	0,86	10	3,5					lager	18	5	
substantiv	0,09	8	efterverkning	0	13	5					havsvåg	0,07	7	
adjektiv	1,71	8	påtaglig	20,57	8	3,5	tydligt	28,29	7					
substantiv	0,86	11	fattigkvarter	0	13	3,7					slum	7,71	4	

id	regel	originalmening	modifierad mening
1	passiv-aktiv	Vägar och järnvägar brukar oftast markeras med färgade linjer av olika slag.	Vägar och järnvägar brukar man oftast markera med färgade linjer av olika slag.
2	passiv-aktiv	Bergarter brukar i vardagligt tal kallas stenar.	Bergarter brukar man i vardagligt tal kalla stenar.
3	passiv-aktiv	Ris kan också odlas på bevattnade terrasser i bergigare områden.	Ris kan man också odla på bevattnade terrasser i bergigare områden.
4	passiv-aktiv	Länet organiserades av medeltidens kungamakt.	Medeltidens kungamakt organiserade länet.
5	passiv-aktiv	I västra Australien bryts järnmalm i enorma dagbrott.	I västra Australien bryter man järnmalm i enorma dagbrott.
6	passiv-aktiv	Proven påstods vara ofarliga.	Man påstod sig att proven kommer vara ofarliga.
7	straight word order	Där breder tundran ut sig med buskar och enstaka låga träd.	Tundran breder där ut sig med buskar och enstaka låga träd.
8	straight word order	Då fungerade allt, minns de, det fanns skola och gratis sjukvård.	Allt fungerade då, minns de, det fanns skola och gratis sjukvård.
9	straight word order	Ibland kan det gå flera år mellan regnen.	Det kan ibland gå flera år mellan regnen.
10	straight word order	Däremot är många salta sjöar rika på fisk.	Många salta sjöar är däremot rika på fisk.
11	straight word order	Då uppstod rymdens minsta beståndsdelar.	Rymdens minsta beståndsdelar uppstod då.
12	split-a	Andra områden, framför allt i norr, är mycket glesbefolkade.	Andra områden, är mycket glesbefolkade, framför allt i norr.
13	split-a	Alla planeterna, utom Merkurius och Venus, har en eller flera månar.	Förutom Merkurius och Venus, så har alla planeterna, en eller flera månar.
14	split-a	Men marken, utom det allra översta lagret, är alltid frusen.	Men marken är alltid frusen, förutom det allra översta lagret.
15	split-r	Bland de länder som har lagt anspråk på en tårtbit är Norge.	Bland de länder är Norge som har lagt anspråk på en tårtbit.
16	split-r	Det är vintermonsunen som börjat blåsa.	Det som börjat blåsa är vintermonsunen.
17	split-r	En cyklist som vilar upp sig i en barrskog.	En cyklist i en barrskog som vilar upp sig.
18	split-k	Skilnaderna är dock stora, både mellan länder och mellan regioner.	Skilnaderna är dock stora mellan länder och mellan regioner.
19	split-k	Rörelserna gör att en del kolliderar, medan andra glider isär.	Rörelserna gör att en del kolliderar, andra glider isär.
#	split-k	Människor är oense om var gränsen ska gå, eller om den ska vara där överhuvudtaget. De är visserligen inte lika mäktiga som Himalaya, men de är ändå mycket svårtillgängliga och har i årtusenden varit besvärliga hinder för människor som måste ta sig fram i trakten.	Människor är oense om var gränsen ska gå. Om den ska vara där överhuvudtaget. De är visserligen inte lika mäktiga som Himalaya. De är ändå mycket svårtillgängliga. De har i årtusenden varit besvärliga hinder för människor som måste ta sig fram i trakten.
21	split-k	Där finns stor tillgång på vatten, och somrarna är varma och soliga.	Där finns stor tillgång på vatten. Somrarna är varma och soliga.
#	split-k	Framsteg har alltså gjorts, men mycket återstår att göra.	Framsteg har alltså gjorts. Men mycket återstår att göra.
#	split-a	De nya spanska regenterna, Ferdinand och Isabella ville inte vara sämre och utrustade en expedition som gjorde att italienaren Christofer Columbus år 1492 kunde korsa Atlanten och nå fram till Västindien.	De nya spanska regenterna, Ferdinand och Isabella ville inte vara sämre och utrustade en expedition. Detta gjorde att italienaren Christofer Columbus år 1492 kunde korsa Atlanten och nå fram till Västindien.
	split-a	Vikingarna handlade med skinn och trälar och de första städerna, Birka och Uppåkra, uppstod på grund av handeln.	Vikingarna handlade med skinn och trälar. De första städerna, Birka och Uppåkra, uppstod på grund av handeln.
	split-a	Kungarna på vikinga- och medeltiden var mycket intresserade av att grunda städer, eftersom det var där som handeln ägde rum och det var enkelt att ta upp skatt, tull, vid stadporten.	Kungarna på vikinga- och medeltiden var mycket intresserade av att grunda städer, eftersom det var där som handeln ägde rum. Det var enkelt att ta upp skatt, tull, vid stadporten.
	split-r	Datasystem, satelliter och Big Data har ökat mängden information som kan användas när dagens informationsrika digitala kartor skapas.	Datasystem, satelliter och big Data har ökat mängden information. Detta kan användas när dagens informationsrika digitala kartor skapas.
	split-r	Genom den nordamerikanska kontinenten går en gräns som är mycket tydlig, nämligen gränsen mellan USA och Mexiko.	Genom den nordamerikanska kontinenten går en gräns. Denna gräns är farlig. Detta är gränsen mellan USA och Mexiko.
	split-r	Men det är ännu bara i Persiska viken som man i stor skala har börjat utnyttja naturresurserna.	Men det är bara i Persiska viken. Där har man i stor skala börjat utnyttja naturresurserna.

id	meningar	omskrivning
1	För att du ska kunna läsa en karta på rätt sätt, måste den ha en nordpil.	För att du ska kunna läsa en karta på rätt sätt, måste den ha en pil som pekar mot norr.
2	I de torra områdena är det svårt att bedriva åkerbruk.	I de torra områdena är det svårt att bruka åkrarna.
3	Men satellitbilder visar att öknen ibland också krymper.	Men bilder från satelliter visar att öknen ibland också krymper.
4	De många bilarna orsakar svåra luftföroreningar.	De många bilarna orsakar svåra föroreningar av luften.
5	Betesdjuren gjorde att vegetationen förändrades.	Djuren som betar gjorde att vegetationen förändrades.

Hus är oftast svarta små fyrkanter, medan kyrkor alltid **symboliseras** med ett kors.

- föreställs
- betecknas**
- Ha kvar **symboliseras**

Click this button to continue

Förklara ordet **beteckna** högt

Click this button to continue

Är det blåa ordet svårt?

Därför finns det inga människor som lever **permanent** på Antarktis.

- Ja
- Nej

Click this button to continue

Vilket ord är lättare än permanent?

- ständigt
- bestående
- inget ord är lättare*

Click this button to continue

Vilket ord tycker du passar in?

På _____ är marken ständigt frusen.

- tundran
- slätten
- hed
- inget ord*

Click this button to continue

Vilket ord är lättast?

- Det är en **formation** som skapas av det rinnande vattnet.
- Det är en **ställning** som skapas av det rinnande vattnet.
- Det är en **klippformation** som skapas av det rinnande vattnet.
- inget ord är lätt*


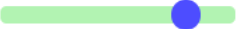

Click this button to continue

Vilken mening är lättast?

- Vägar och järnvägar brukar oftast markeras med färgade linjer av olika slag.
- Vägar och järnvägar brukar man oftast markera med färgade linjer av olika slag.
- Spelar ingen roll

Click this button to continue

Hur lätt är meningen:
"Bergarter brukar i vardagligt tal kallas stenar"?

 (lite)   (mycket)

6

Click this button to continue

Vad tyckte du om meningen
"Bergarter brukar i vardagligt tal kallas stenar"?
(prata högt)

Click this button to continue

Vilket alternativ är tydligast?

- För att du ska kunna läsa en karta på rätt sätt, måste den ha en **nordpil**.
- För att du ska kunna läsa en karta på rätt sätt, måste den ha en **pil som pekar mot norr**.
- Spelar ingen roll

Click this button to continue

Förklara ordet: "nordpil" (prata högt)

Click this button to continue

Appendix 3

Vilket kön tillhör du?

- Man
- Kvinna
- Annat
- Vill inte svara

Click this button to continue

hur gammal är du?

Skriv ner din ålder

Click this button to continue

Vilken skola går du i:

yll i här

Click this button to continue

Vilken nivå pluggar du på?

- Grundskolan
- Gymnasiet
- Högskola/universitet
- Inget av alternativen

Click this button to continue

Fråga testledaren vad du ska skriva in här

yll i här

Click this button to continue

har du fått information om testet och samtycke, samt skrivit under en samtyckesblankett?

- Ja
- Nej

Click this button to continue

Vilket ord tycker du passar in?

Varför har framstegen kommit under vissa _____ tidsperioder?

- särskilda
- speciella
- bestämda
- inget ord*

Click this button to continue

Varför valde du ordet **särskilda**?
Varför har framstegen kommit under vissa särskilda tidsperioder?

Click this button to continue

Vilket ord är lättast?

- Dessutom kan det komma **nederbörd** när som helst.
- Dessutom kan det komma **regn** när som helst.
- Inget ord är lätt*

Click this button to continue

Varför valde du ordet **nederbörd**?
Dessutom kan det komma nederbörd när som helst.

Click this button to continue

Vilken mening är lättast?

- Rymdens minsta beståndsdelar uppstod då.
- Då uppstod rymdens minsta beståndsdelar.
- Spelar ingen roll**

Click this button to continue

Tycker du meningen är lätt?
**"Rymdens minsta beståndsdelar
uppstod då"**

- Ja
- Nej

Click this button to continue

Varför valde du meningen:
**"Rymdens minsta beståndsdelar
uppstod då"?**
(prata högt)

Click this button to continue

Vilken är lättast?

- I de torra områdena är det svårt att bedriva **återbruk**.
- I de torra områdena är det svårt att **bruka åkrarna**.
- Spelar ingen roll**

Click this button to continue

Varför valde du **"bruka åkrarna"**?
(prata högt)

Click this button to continue

