

YDS AS A BENCHMARK IN TURKEY: THE DIFFICULTY LEVELS OF READING COMPREHENSION QUESTIONS

Cüneyt DEMİR^(*)

Abstract

The ability to speak a language covers all skills, namely reading, writing, speaking, and listening, together with a reasonable degree of performance and competence knowledge. Specializing over a single skill by overlooking the others may hinder speaker proficiency. In other words, all skills equally weigh; hence, the tests that aim to measure language proficiency should be able to gauge all skills in a good balance. In spite of that reality, YDS (Foreign Language Exam) in Turkey scores test-takers only through reading questions, and accordingly, this test may not be a good benchmark for Turkey because reading questions do not seem to have a stable difficulty level. The exam is composed of different kinds of questions which are vocabulary, grammar, comprehension, dialogue, and reading. The reading questions take up the biggest portion in the exam -20 out of 80 questions-; therefore, they highly determine the test-takers' scores. This study investigated eleven YDS tests held in the last 5 years. In total eleven YDS tests (220 reading questions) were analysed. Coleman-Liau Index was used to measure how difficult each test is. This study aims to reveal whether reading questions of YDS have a similar difficulty level by calculating the Coleman-Liau Index. This study is of great importance because YDS is a crucial benchmark test in Turkey that determines the English proficiency level of those who take it, and any discrepancy in difficulty levels may negatively affect the reliability of the test. The results showed that the reading questions are in a large range of difficulty, and do not follow a stable track of difficulty.

Keywords: YDS, Benchmark, Difficulty, Reading, Coleman.

^{*)} Dr. Öğr. Üyesi, Yabancı Diller Yüksekokulu Mütercim Tercümanlık (İngilizce)
(e-posta: cuneytdemir@siirt.edu.tr) ORCID ID: <https://orcid.org/0000-0003-2588-372X>

Türkiye’de Yabancı Dil Becerisini Ölçme Sınavı Olarak YDS: Okuduğunu Anlama Sorularının Zorluk Düzeylerinin İncelenmesi

Öz

Bir dili kullanabilme, onun okuma, yazma, konuşma ve dinleme gibi bütün becerileriyle kullanabilmeyi gerektirmektedir. Bir beceri üzerinde gelişirken bir diğ erinin zayıf bırakılması dili öğrenen kişinin tam akıcı olması önünde bir engeldir. Diğ er bir ifadeyle, bir dili ölçerken bütün becerileriyle ölçmek gereklidir. Bu gerçeğe rağmen Türkiye’de uygulanan YDS sadece okuma becerisi üzerinden puanlama yapmaktadır. Fakat ne yazık ki bu test standart bir zorluk seviyesine sahip olmadığı için iyi bir ölçme sınavı olmayabilir. Bu sınav kelime bilgisi, dil bilgisi, okuduğunu anlama, diyalog ve okuma sorularından oluşmaktadır. Seksen sorunun yirmisini oluşturan okuma soruları sınavda en fazla ağırlığı olan soru türüdür ve bundan dolayı puanlamada esas teşkil eder. Bu çalışma YDS’de son 5 yıl içerisinde sorulan toplam on bir sınavın okuduğunu anlama sorularını analiz etti. Toplamda 220 okuduğunu anlama sorusu analiz edildi. Her testin hangi zorluk düzeyinde olduğunu bulmak için Coleman-Liau analiz tekniğini kullandı. Bu çalışma Coleman-Liau tekniğiyle YDS sınavında okuduğunu anlama sorularının aynı zorluk düzeyinde olup olmadıklarını araştırmayı amaçlamaktadır. YDS Türkiye’de yabancı dil becerisini ölçen en önemli sınav olduğu için, bu çalışma testin güvenilirliği konusunda bir fikir edinmede büyük bir önem arz etmektedir ve soruların zorluk düzeylerinde ki büyük değişiklikler sınav güvenilirliğini etkileyecektir. Sonuçlar okuma parçalarının zorluk düzeyleri arasında farklılıkların olduğunu ve standart bir zorluk düzeyine sahip olmadıklarını göstermiştir.

Anahtar Kelimeler: YDS, Ölçüt, Zorluk, Okuma, Coleman.

1. Introduction

The purpose of an exam is to measure knowledge. Different from classical exams, language exams measure not only the knowledge but also the proficiency, and while doing this, the questions, either written or spoken, are prepared meticulously so that the reliability of the exams would not be negatively affected. Language testing has been gaining popularity all over the world and countries make their exams so that they can determine the language proficiencies of their citizens. For example, Canada has its language testing of CELPIP (The Canadian English Language Proficiency Index Program); America has TOEFL; UK has IELTS; Turkey has YDS and so on. Of all international English language tests, TOEFL and IELTS are the most prevalent ones; valid nearly in all countries. A common feature of these exams is that they measure not a single skill but a general proficiency of the speaker such as writing, reading, listening and speaking. Compared to language tests gauging general language skills, YDS does not test the general language proficiency of a speaker but only vocabulary, grammar and reading, and all measuring require you to solve multiple-choice questions.

There are many motives for taking language proficiency tests such as qualification for a position and acceptance for a university. Whatever the reason is, many people struggle with tests to have a good mark, and take the test repeatedly particularly when they get fail, or are not able to get the required score they need. There are occasions when a test taker was not able to pass in the first exam while s/he did in the subsequent one, which should be because the test-taker studied more; however, the case is not always like this, and the test taker pass just because the difficulty level of the test was reduced. This is an infelicity in terms of test reliability, and must be dealt with. Language proficiency testing is particularly sensible to test reliability as any meaningless fluctuations in difficulty may result in unfair scores, hence inaccurate assessment of test-takers.

Assessment is crucial to determine the language levels of speakers and has a significant role in the process of teaching and learning (Lam, 2015). Traditional human-based assessment is affected by different factors such as context, psychological state of the test taker, assessor and motivation; therefore, anything that may affect one of these factors may prompt wrong-assessment, and accordingly scoring error. However, the standard quality of a benchmark is the most important factor particularly if it is implemented on large scales. Any fluctuation in difficulty may result in insure scoring, i.e. the levels of test-takers cannot be measured fairly. In line with this, this study aims to analyse reading comprehension questions in YDS (Foreign Language exam) between 2013 and 2018 to track the difficulty levels of the exams. This study proposes that the stability of readability of comprehension questions in YDS should be maintained to acquire reliable test results. Only reading comprehension questions were evaluated because reading comprehension is a complex and multidimensional process (Carlson, Seipel, & McMaster, 2014); therefore it can readily measure English proficiency

1.1. What is YDS?

YDS is a test held in different languages to measure language proficiency in Turkey. The acronym of YDS, called KPDS before 2013, stands for *Yabancı Dil Testi* in Turkish, the English of which would be Foreign Language Test. It was being. YDS is held biannually and has 80 questions. It is not a test measuring general English skills but just reading skill. The questions are composed of vocabulary (6), grammar (10), cloze tests (10), sentence completion (10), translation (6), reading comprehension (20), dialogue (5), closest in meaning (4), paragraph completion (4), and irrelevant sentence (5). You need to take this test to continue your higher education, to be an academician or for other specific purposes. A test-taker should have a score of at least 50 to be successful. The scoring is done on a scale of 100. The equivalent qualifications of YDS are provided in table 1.

Table 1. Equivalent Qualifications and Tests

YDS	TOEFL iBT	PTE Academic	CPE	CAE
100	120	90	A	
90	108	84	B	
80	96	78	C	A
70	84	71		B
60	72	55		C
50	60	45		

1.2. Reading Comprehension Questions and Scoring

Compared to other types of questions, reading comprehension questions requires multiple language proficiencies such as not only knowledge of word recognition skills but also other linguistic and extralinguistic skills. This multidimensional domain embodies a number of language components that necessitate readers to have a competent aptitude for language proficiency. Therefore, apart from fluency in word recognition (Perfetti, 1999), the reader is to be capable of analysing morphological awareness (Deacon, Kieffer, & Laroche, 2014), metacognitive knowledge (Perfetti, Landi, & Oakhill, 2005), and syntactic awareness (Nagy, 2007). Accordingly, only reading comprehension questions were evaluated in this study mainly for two reasons: the first, they constitute the majority of the questions in YDS; therefore, they are the most influential part of scoring. The second, reading comprehension is a unitary skill, i.e. reading comprehension is a complex and multidimensional process (Carlson, Seipel, & McMaster, 2014); therefore, in order to decode the texts, a test-taker needs to have a competence in grammar, a high lexical reservoir, and pragmatic knowledge of the context. Concisely, properly prepared reading comprehension questions are more qualified to gauge the English proficiency of test-takers than any other questions types, which is why reading comprehension questions is of paramount importance for any test aiming at measuring language skill.

1.3. Research Aim and Research Questions

The present study aims to measure the difficulty level of reading comprehension questions of YDS implemented between 2013 and 2018 by using Coleman-Liau index, and to show the most difficult and the easiest reading passages (research question 1). A common view with comprehension questions in Turkey is that they necessitate a large lexical reservoir, hence are difficult when compared to other types of questions. This study aims to investigate whether there is a positive or negative correlation between CLI and mean exam successes of test-takers (research question 2). Finally, each YDS has five reading passages and the calibration of these passages is important; it is not

reliable for an exam to have reading passages at different difficulties. In line with that, this study aims to investigate which YDS has the most volatility in the difficulty of reading comprehension passages (research question 3). Specifically, this study aims to answer the research questions below.

1. What term of YDS has the highest and the lowest CLI regarding reading comprehension passages?
2. Is there any correlation between CLI and mean exam success?
3. What term of YDS has the most and least volatility in difficulty of reading comprehension passages?

2. Methodology

2.1. Data

The data are composed of eleven YDS exams held between only 2013-2018 years because the exams had been held under a different name before 2013; namely KPDS. Furthermore, the number of questions has been decreased to 80 from 100 in the wake of KPDS, and also the number of reading comprehension questions. Therefore, the exams before YDS, which were KPDS, were excluded from the data. Each YDS exam includes five reading passages and four questions for each passage; in total twenty reading comprehension questions. Table 2 provides detailed information about the questions analysed.

Table 2. The Detailed Content of the Data

Exam	Number of Passages	Number of questions	Number of words
2018 Spring Term	5	20	1012
2017 Fall Term	5	20	1058
2017 Spring Term	5	20	1105
2016 Fall Term	5	20	1060
2016 Spring Term	5	20	1029
2015 Fall Term	5	20	982
2015 Spring Term	5	20	1116
2014 Fall Term	5	20	970
2014 Spring Term	5	20	1020
2013 Fall Term	5	20	1024
2013 Spring Term	5	20	1005
TOTAL	55	220	11381 (1035 \bar{x})

2.2. Missing Data

Two exams are held each year; spring term and fall term. In spite of this, the data included only a single exam in 2018 (spring term) because the institution holding YDS did not share the exam although it was contacted to them. Additionally, this study did not analyse all question types in YDS but only those of reading comprehension questions because readability tests such as CLI need a certain number of words to be able to calculate the difficulty level. The other questions such as vocabulary, grammar, and so forth do not have enough words to reach a reliable ramification. For this reason, the present study only analysed reading comprehension questions; in addition, they constitute a large part of the whole exam when compared to other question types.

2.3. Coleman-Liau index (CLI)

This study used Coleman-Liau index (CLI) to investigate the difficulty level of each exam. CLI is a readability test that investigates how difficult a text is, thus allows us to understand the grade level. It is widely used in the USA and some other countries. Unlike most of the other grade-level predictors, CLI depends on characters instead of syllables per word. In other words, instead of referring to syllable/word and sentence length indices, “Meri Coleman and T. L. Liau believed that computerized assessments understand characters more easily and accurately than counting syllables and sentence length” (Readable Blog, 2017); therefore, this index is superior to other readability indices relying on syllable-counting techniques. The formula that CLI used is shown in figure 1. The summary of the formula is this: $(5.89 * \text{characters/words}) - (0.3 * \text{sentences} / (100 * \text{words})) - 15.8$.

$$CLI = 5.89 \left(\frac{\text{characters}}{\text{words}} \right) - 0.3 \left(\frac{\text{sentences}}{100 \text{ words}} \right) - 15.8$$

Figure 1. The formula of coleman-liau index.

The high number of CLI should be interpreted as a difficult reading passage while lower CLI figures show the easiness of the passages. The number you obtain is the grading equivalence of the results; for example, CLI of 12 refers to grade 12 (college level) or 4 refers to grade 4 (primary school level).

2.4. Procedure

The exams are legally available on the web page of the institution implementing the YDS exams (OSYM, 2016). Having compiled the exams, the researcher picked up reading comprehension questions. A total fifty-five reading passages were collected and each reading passage was analysed through the formula in figure 1. A second assessor, a

specialist on statistics, called for calculating CLI of randomly selected six passages (11% of whole data) in case the researcher might make miscalculation, and found the same results as me. The CLI of each passage and each exam was delivered through tables in the result section. Each table presents the number of characters, number of words, number of sentences, the average number of characters per word, the average number of syllables per word, and the average number of words per sentence because all these data are necessary to make accurate calculations they are important in calculating the CLI accurately.

3. Results

Results were provided through tables, each referring to a different exam. The analyses started from 2013 fall term YDS exam to 2018 spring term exam. Table 3 shows the results regarding YDS 2013 fall term.

Table 3. CLI Analysis Results of YDS 2013 Fall Term Exam.

YDS 2013 Fall Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1058	1058	1089	1033	1023	1052
Number of words	202	216	229	192	203	208
Number of sentences	10	9	9	8	13	10
Average number of characters per word	5.24	4.90	4.76	5.38	5.04	5
Average number of syllables per word	1.80	1.68	1.58	1.86	1.76	1.7
Average number of words per sentence	20.20	24	25.44	24	15.62	21.3
CLI	13.56	11.80	11.03	14.64	11.96	12.53

The CLI differences between passages are not marked and relatively close to one another, which represents a smooth transition between passages. The 4th passage has the highest CLI while the 3rd has the lowest, and the mean CLI for this exam is 12.53. Table 4 shows results regarding 2013 YDS spring term.

Table 4. CLI Analysis Results of YDS 2013 Spring Term Exam.

YDS 2013 Spring Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1049	990	1156	999	1011	1041
Number of words	187	194	219	216	202	204
Number of sentences	8	8	11	13	8	9.6
Average number of characters per word	5.61	5.1	5.3	4.6	5	5.1
Average number of syllables per word	1.94	1.8	1.8	1.6	1.7	1.8
Average number of words per sentence	23.4	24.3	20	16.7	25.3	21.2
CLI	15.96	13	13.78	9.64	12.49	12.90

Different from the former exam, this exam has a high CLI volatility. Furthermore, mean CLI is one of the highest compared to other exams; 12.90. The range between the passage with the highest CLI and the lowest is 6.32, respectively the first and the fourth passages. Table 5 shows the results concerning 2014 YDS fall term.

Table 5. CLI Analysis Results of YDS 2014 Fall Term Exam.

YDS 2014 Fall Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1086	1096	1000	851	1051	1017
Number of words	221	192	204	169	200	197
Number of sentences	9	9	13	10	11	10.4
Average number of characters per word	4.9	5.7	4.9	5	5.3	5.17
Average number of syllables per word	1.7	1.9	1.7	1.7	1.8	1.7
Average number of words per sentence	24.6	21.3	15.7	16.9	18.2	19
CLI	11.92	16.42	11.16	12.08	13.50	12.99

Similarly, this exam has high volatility; it is not as high as the former, though. The most difficult passage in this exam is the second one whereas the easiest is the first one. On the other hand, the exam has one of the highest mean CLI; 12.99. Results of 2014 spring term were provided in table 6.

Table 6. CLI Analysis Results of YDS 2014 Spring Term Exam.

YDS 2014 Spring Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1123	1114	1071	985	1010	1060
Number of words	220	190	214	186	226	207
Number of sentences	11	9	9	7	11	9.4
Average number of characters per word	5.1	5.9	5	5.3	4.5	5.1
Average number of syllables per word	1.7	2	1.7	1.8	1.6	1.8
Average number of words per sentence	20	21.1	23.8	26.6	21	22
CLI	12.77	17.31	12.42	14.26	9.06	12.99

Passage two has the highest CLI while passage five has the lowest. The average CLI of this exam is 12.99. Similar to the fall term exam of the same year, this study also has high volatility in the difficulty of reading passages. Table 7 provides the details of 2015 YDS fall term.

Table 7. CLI Analysis Results of YDS 2015 Fall Term Exam.

YDS 2015 Fall Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	982	998	1011	1029	928	990
Number of words	202	198	190	219	188	199
Number of sentences	7	8	10	9	9	8.6
Average number of characters per word	4.9	5	5.3	4.7	4.9	45
Average number of syllables per word	1.6	1.7	1.8	1.6	1.7	1.7
Average number of words per sentence	29	24.8	19	24.3	21	23.1
CLI	12.09	12.68	13.96	10.64	11.84	12.20

It is understood that the third and fifth passages are the easiest and hardest ones, respectively. The mean CLI is 12.20 for this exam. This study has smooth volatility in difficult. Table 8 provides the results of 2015 YDS spring term.

Table 8. CLI Analysis Results of YDS 2015 Spring Term Exam.

YDS 2015 Spring Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1206	1113	976	1129	1105	1106
Number of words	254	233	195	222	229	227
Number of sentences	14	11	8	10	11	10.8
Average number of characters per word	4.8	4.8	5	5.1	4.8	4.9
Average number of syllables per word	1.6	1.6	1.7	1.7	1.7	1.7
Average number of words per sentence	18.1	21.2	24.4	22.2	20.8	21
CLI	10.51	10.92	12.45	12.80	11.18	11.51

The passage with the lowest CLI is the first one while the fourth one had the highest CLI. The average CLI for this exam is 11.51. Smooth volatility in difficulty exists. Table 9 shows the details of 2016 YDS fall term.

Table 9. CLI Analysis Results of YDS 2016 Fall Term Exam.

YDS 2016 Fall Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1063	1110	961	955	1178	1053
Number of words	238	200	192	202	239	214
Number of sentences	11	10	11	12	13	11.4
Average number of characters per word	4.5	5.6	5	4.7	4.9	4.9
Average number of syllables per word	1.5	1.8	1.6	1.5	1.7	1.6
Average number of words per sentence	21.7	20	17.5	16.7	18.4	18.8
CLI	9.12	15.39	11.96	10.26	11.60	11.57

The results show volatility in the difficult of reading passages of 2016 YDS fall term. The lowest CLI is 9.12 (passage 1) while the highest is 15.39 (second passage). The mean CLI is 9.12. Table 10 informs about the results of 2016 YDS spring term.

Table 10. CLI Analysis Results of YDS 2016 Spring Term Exam.

YDS 2016 Spring Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	965	1113	1121	843	837	976
Number of words	193	234	220	196	191	207
Number of sentences	8	15	10	12	9	10.8
Average number of characters per word	5	4.8	5.1	4.3	4.4	4.7
Average number of syllables per word	1.7	1.5	1.8	1.4	1.5	1.6
Average number of words per sentence	24.1	15.6	22	16.3	21.2	19.2
CLI	12.41	10.29	12.85	7.7	8.60	10.43

To the results, the average CLI is 10.43. Passages four and three have respectively the lowest and highest CLI, 7.7, 12.85. 10.43 is the mean CLI. Table 11 presents the results of 2017 YDS fall term.

Table 11. CLI Analysis Results of YDS 2017 Fall Term Exam.

YDS 2017 Fall Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	983	1050	1110	996	1047	1037
Number of words	208	231	221	197	215	214
Number of sentences	16	16	10	9	10	12.2
Average number of characters per word	4.7	4.6	5	5	4.9	4.8
Average number of syllables per word	1.6	1.5	1.7	1.8	1.7	1.6
Average number of words per sentence	13	14.4	22.1	21.9	21.5	17.6
CLI	9.73	8.89	12.43	12.61	11.49	10.99

There is smooth CLI difference among passages in this exam. The lowest CLI belongs to passage two while the highest belongs to passage four. The mean CLI is 10.99. The next table shows the details of 2017 YDS spring term.

Table 12. CLI Analysis Results of YDS 2017 Spring Term Exam.

YDS 2017 Spring Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	1026	1063	1158	1178	1005	1086
Number of words	210	216	236	249	203	223
Number of sentences	12	11	14	10	10	11.4
Average number of characters per word	4.9	4.9	4.9	4.7	4.9	4.9
Average number of syllables per word	1.6	1.6	1.7	1.6	1.7	1.6
Average number of words per sentence	17.5	19.6	16.9	24.9	20.3	19.6
CLI	11.26	11.66	11.32	10.86	11.88	11.37

There is a perfect harmony of CLI with this exam. Almost no volatility was detected. Passages with the highest and the lowest CLI is respectively, 11.88 and 10.86. The average CLI is 11.37. Table 13 provides the details of the last exam: 2018 YDS spring term.

Table 13. CLI Analysis Results of YDS 2018 Spring Term Exam.

YDS 2018 Spring Term	Passage 1	Passage 2	Passage 3	Passage 4	Passage 5	Average
Number of characters	725	971	1323	1029	801	970
Number of words	154	220	270	219	166	206
Number of sentences	8	11	12	8	10	9.8
Average number of characters per word	4.7	4.4	4.9	4.7	4.8	4.7
Average number of syllables per word	1.6	1.5	1.6	1.6	1.6	1.6
Average number of words per sentence	19.3	20	22.5	27.4	16.6	21
CLI	10.37	8.70	11.73	10.78	10.81	10.53

There is even volatility with this exam. 11.73 (passage 3) is the highest CLI and 8.70 (passage two) is the lowest CLI. The mean CLI is 10.53.

Table 14. Mean CLI and Mean Success Rate of all Exams in the Data.

Exam	Number of words	Number of sentences	CLI	Mean success rate*
2013 Fall	1042	49	12.53	36.76
2013 Spring	1018	48	12.90	30.47
2014 Fall	986	52	12.99	41.37
2014 Spring	1036	47	12.99	31.62
2015 Fall	997	43	12.20	37.7
2015 Spring	1133	54	11.51	34.1
2016 Fall	1071	57	11.57	38.65
2016 Spring	1034	54	10.43	32
2017 Fall	1072	61	10.99	39.46
2017 Spring	1114	57	11.37	36
2018 Spring	1029	49	10.53	37.75

* The mean success rate scores are delivered out of 100.

To the table, the exam with the highest mean success rate is 2014 fall term with a score of 41.37, and then 2017 YDS fall term (39.46) and 2016 fall term (38.65) follow. Although they have an average of the highest mean success rates, they do not have the lowest CLI. The exams with the highest CLI are 2013 spring term, 2014 fall term, and 2014 spring term. Table 15 presents the summary of CLI differences between passages.

Table 15. Summary of CLI of Each Exam.

Exam	Highest CLI	Lowest CLI	Difference
2013 Fall	14.64	11.03	3.61
2013 Spring	15.96	9.64	6.32
2014 Fall	16.42	11.16	5.26
2014 Spring	17.31	9.06	8.25
2015 Fall	13.96	10.64	3.32
2015 Spring	12.80	10.51	2.29
2016 Fall	15.39	9.12	6.27
2016 Spring	12.85	7.7	5.15
2017 Fall	12.61	8.89	3.72
2017 Spring	11.88	10.86	1.02
2018 Spring	11.73	8.70	3.03

As table 15 shows, the difference of CLI between the passages is lowest in 2017 YDS spring term -1.02- while it is highest in 2014 YDS spring term, 8.25.

4. Discussion

The present study aimed to measure the difficulty level of reading comprehension questions of YDS implemented between 2013 and 2018 by using Coleman-Liau index; to show the most difficult and the easiest reading passages; to investigate whether there is a positive or negative correlation between CLI and mean exam successes of test-takers, and to investigate which YDS has the most volatility in difficulty of reading comprehension passages. As to these purposes, this study answered the research questions as follows.

1. What term of YDS has the highest and the lowest CLI regarding reading comprehension passages?

The results showed that both 2014 spring and fall terms YDS have the highest CLI. CLI refers to the difficulty of reading passages and because these exams have the highest reading passages, it can be concluded that 2014 YDS spring and fall terms are the exams with the most difficult passages. On the other hand, 2016 YDS spring term came forward as the exam with the lowest CLI, which means that it is the exam with the easiest reading passages.

On the other hand, this difference of CLI is not appropriate for a qualified test because the test should follow a straight line in difficulty so that the test-takers would not feel anxiety about the difficulty level of reading questions that take up the big portion in scoring. The questions can be ordered from easy to difficult (Zimmaro, 2016) but this has nothing with ordering passages in the same way. Ensuring accurate difficulty level of questions is essential to help test-takers learn more efficiently and effectively (Pérez, Santos, Pérez, de Castro Fernández, & Martín, 2012) for the prospective exams.

2. Is there any correlation between CLI and mean exam success?

This is a crucial but difficult question to answer because you need to make thorough analyses to reach a stable conclusion. However, the question can be answered partly in terms of investigating the relationship between the mean scores and mean CLI. According to the results, there is not a direct positive relation between CLI and success rate. For example, 2014 YDS spring and fall terms exams have the most difficult reading passages of all other exams while the mean scores are not the lowest. In other words, although the CLI is 12.99 in YDS 2014 fall term (the highest CLI), the mean success rate is the highest of all exams (see table 14). Again, although the CLI is 12.99 in YDS 2014 spring term, the mean success rate is the second-worst. This shows that there is not a direct relationship between CLI and success rate, but this is not to say that there is an

asymmetrical relationship because results show that there are exams in which both CLI and mean success rate is low (table 14)

3. What term of YDS has the most and least volatility in difficulty of reading comprehension passages?

The degree of difficulty may change from one exam to another; however, the change of difficulty may occur inside a single exam. This study found that each exam does not have an even degree of difficulty in reading passages (see table 15). For example, the CLI difference among the passages is 8.25 in 2014 YDS fall term exam, which means that one passage is very easy while the other is very difficult. Although it was observed that this difference exists in the majority of the exams, some exams, 2015 and 2017 YDS spring terms, have smooth volatility among passages.

Comprehending a text is a sophisticated process depending on the models of reading comprehension (Trapman, Gelderen, Schooten, & Hulstijn, 2016) because readers would need to apply their linguistic knowledge as well as appropriate mental representation of a text (Perfetti, Landi, & Oakhill, The acquisition of reading comprehension skills, 2005). Therefore, providing passages at similar difficulty is necessary to avoid possible confusion with the test-takers. A reading passage composed of simple or difficult words would determine the comprehension level of the test-taker (Susanti, Nishikawa, Tokunaga, & Hiroyuki, 2016). However, a smooth transition in degree of difficulty is an obligation for test reliability; therefore, “caution should be exercised when selecting tests for the assessment of reading difficulties (Nation, 1997, p. 359)” particularly for the tests such as YDS that is composed of more sophisticated and compound sentences (Aşkaroğlu, 2013).

Individual differences of test-takers in reading comprehension questions are attributed to the knowledge of linguistic knowledge particularly vocabulary and grammar (Van Gelderen, et al., 2004); therefore, decoding fluency is of paramount importance to define the level of reading comprehension, and accordingly irregular fluctuations between paragraphs may negatively affect the flair for decoding fluency i.e. speed of sentence processing which is one of the crucial factors that determines reading comprehension ability (Van Gelderen, Schoonen, Stoel, De Gloppe, & Hulstijn, 2007). On the other hand, carefully planned reading comprehension paragraphs may have a positive washback effect on the reading skills of test-takers, as was evidenced by Akpınar and Çakildere (2013) who also investigated YDS.

Another caveat about tests is that whether they are for level-grading or to measure proficiency. Level-graded questions may be included in a test to determine the levels of the test-takers, and accordingly to rank them as low-achieving or high-achieving language speakers. However, high-stakes tests such as YDS -considered to be proficiency tests-address to a group with high language aptitude and are carried out to be used in areas such as academia that necessitate high level of language proficiency, which is why tests as

YDS is to have a stable level of difficulty. Given the evidence this study concluded, some tests (table 15) had CLI over normalised levels in terms of paragraph difficulty whereas some others had only minor differences. Besides, major CLI fluctuations exist among tests that were set at different terms, which may be due to their paragraph style. Kıray (2015) stated that there is not a standardized paragraph style in YDS exams; in other words, some paragraphs seem to be deficient because they are not full but a small fraction of a whole text. Therefore, simply picking a paragraph out of a full text may lead to difficulty at various levels, if not ambiguity in terms of reading comprehension. Finally, the results are in line with the assumptions of Gur (Gür, 2013) who indicated that KPDS (former name of YDS) did not represent for stable paragraph orders, i.e. some paragraphs are much more difficult than the others.

5. Conclusion

The function of performing an exam is to enable instructors to make judgments about the quality of learning (Piontek, 2008) and to assess the level of preparation reached by the test-takers (Frosini, Lazzarini, & Marcelloni, 1998). Accordingly, reading comprehension questions out of all other question types are the most important one to measure the English proficiency of students, and for the success in every academic domain (Wigfield, Gladstone, & Turci, 2016). Therefore, accurate question preparing for reading comprehension is obligatory for testing-institutions.

Given the importance of assessment, testing institutions should develop exams that include questions balanced in difficulty. If creating high-quality educational assessments is desired, striking a balance on the difficulty level of questions will increase the reliability of exam results. Any fluctuation in difficulty throughout the exams may prevent assessors from reaching a stable assessment on the proficiency of the test-takers because unbalanced exams with questions at different difficulty levels may prompt test-takers to do low-performing and high-performing although they have similar language proficiencies. In brief, difficulty can affect the tests negatively, which is why it should be considered on the process of preparing questions (Tan & Othman, 2013). Accordingly, this study emphasizes that the calibration of reading passages is of importance for both reliable results and accurate measuring. Testing institutions may use different methods -instead of CLI- such as Flesch Kincaid, SMOG, ARI to calculate difficulty level of reading passages.

This study measured the difficulty levels of only reading passages asked in YDS between 2013 and 2018. Further studies may measure the whole questions asked in the questions. Also, tests include vocabulary questions. They are also important in determining the proficiency levels; however, some testing institutions ignore creating equilibrium on the difficulty levels of the vocabularies. Some words may be at A2 level while some others at C1 level. Researchers are kindly invited to investigate this unstable situation because it may have a profound impact on test reliability. Though they are still developing, there are some PC-based programs such as *Collins* providing a degree of difficulty of vocabularies.

References

- Akpınar, K. D., & Cakildere, B. (2013). Washback effects of high-stakes language tests of Turkey (KPDS and ÜDS) on productive and receptive skills of academic personnel. *Journal of Language and Linguistic Studies*, 9(2), 81-94.
- Aşkaroğlu, V. (2013). 2013 YDS ile 2012 KPDS ve ÜDS Sınavlarının Karşılaştırılması. *Karadeniz Uluslararası Dergi*, 162-173.
- Carlson, S. E., Seipel, B., & McMaster, K. (2014). Development of a new reading comprehension assessment: Identifying comprehension differences among readers. *Learning and Individual Differences*, 31, 40-53.
- Deacon, S. H., Kieffer, M. J., & Laroche, A. (2014). The relation between morphological awareness and reading comprehension: Evidence from mediation and longitudinal models. *Scientific Studies of Reading*, 18, 432-451, DOI:10.1080/10888438.2014.926907.
- Frosini, G., Lazzarini, B., & Marcelloni, F. (1998). Performing automatic exams. *Computers & Education*, 281-300.
- Gür, Ö. (2013). Ölçme, değerlendirme ve kamu personeli dil sınavı (KPDS) – Bu sınav neyi ölçüyor? *Sakarya University Journal of Education*, 23-32.
- Kıray, G. (2015). Macro-structure Analysis of Reading Comprehension Paragraphs of KPDS and YDS Exams within Years 2003-2013. *Hasan Ali Yücel Eğitim Fakültesi Dergisi*, 219-233.
- Lam, R. (2015). Language assessment training in Hong Kong: Implications for language assessment literacy. *Language Testing*, 32(2), 169-197.
- Nagy, W. E. (2007). Metalinguistic awareness and the vocabulary-comprehension connection. In R. K. Wagner, A. E. Muse, & K. R. (Eds.), *Vocabulary acquisition: Implications for reading comprehension* (pp. 52-77). New York: Guilford Press.
- Nation, K. (1997). Assessing reading difficulties: the validity and utility of current measures of reading skill. *British Journal of Educational Psychology*, 67(3): 359-370 doi.org/10.1111/j.2044-8279.1997.tb01250.x.
- OSYM. (2016). www.osym.gov.tr. Retrieved 23, 2019, from YDS Çıkmış Sorular: <http://www.osym.gov.tr/TR,15073/yds-cikmis-sorular.html>
- Pérez, E. V., Santos, L. M., Pérez, M. J., de Castro Fernández, J. P., & Martín, R. G. (2012). Automatic Classification of Question Difficulty Automatic Classification of Question Difficulty. *Frontiers in Education Conference* (pp. 1-5). IEEE.
- Perfetti, C. A. (1999). Comprehending written language: A blueprint of the reader. In C. M. Brown, & P. (. Hagoort, *The neurocognition of language* (pp. 167-208). Oxford: Oxford University Press.

- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skills. In C. H. M. J. Snowling, *The science of reading: A handbook* (pp. 227-247). London: Blackwell.
- Perfetti, C. A., Landi, N., & Oakhill, J. (2005). The acquisition of reading comprehension skills. In M. J. Snowling, & C. H. (Eds.), *The science of reading: A handbook* (pp. 227-247). London: Blackwell.
- Piontek, M. E. (2008). Best practices for designing and grading exams. *Occasional Paper*, 24, 1-12.
- Readable Blog*. (2017, 5 22). Retrieved 1 2, 2019, from Readability and the Coleman-Liau Index: <https://readable.com/blog/the-coleman-liau-index/>
- Susanti, Y., Nishikawa, H., Tokunaga, T., & Hiroyuki, O. (2016). Item difficulty analysis of english vocabulary questions. *The 8th International Conference on Computer Supported Education*, pp. 267-274.
- Tan, Y. T., & Othman, A. R. (2013). The relationship between complexity (taxonomy) and difficulty. *The 20th National Symposium on Mathematical Sciences* (pp. 596-603). AIP Publishing LLC.
- Trapman, M., Gelderen, A., Schooten, E., & Hulstijn, J. (2016). Reading Comprehension Level and Development in Native and Language Minority Adolescent Low Achievers: Roles of Linguistic and Metacognitive Knowledge and Fluency. *Reading & Writing Quarterly*, 239-257, DOI: 10.1080/10573569.2016.1183541.
- Van Gelderen, A., Schoonen, R., De Glopper, K., Hulstijn, J., Snellings, P., Simis, A., et al. (2004). Linguistic knowledge, processing speed and metacognitive knowledge in first and second language reading comprehension: A componential analysis. *Journal of Educational Psychology*, 96, 19–30. DOI:10.1037/0022-0663.96.1.19.
- Van Gelderen, A., Schoonen, R., Stoel, R. D., De Glopper, K., & Hulstijn, J. (2007). Development of adolescent reading comprehension in language 1 and language 2: A longitudinal analysis of constituent components. *Journal of Educational Psychology*, 99, 477–491. DOI:10.1037/0022-0663.99.3.477.
- Wigfield, A., Gladstone, J. R., & Turci, L. (2016). Beyond cognition: Reading motivation and reading comprehension. *Child development perspectives*, 10(3), 190-195.
- Zimmaro, D. M. (2016, 11 1). *Writing Good Multiple-Choice Exams*. Retrieved 2 10, 2019, from <https://facultyinnovate.utexas.edu/sites/default/files/writing-good-multiple-choice-exams-fig-120116.pdf>