

Development of an A2 Level Grammar Achievement Test in Accordance with CEFR

Ahmet KÜÇÜKTEKELİ

*Balıkesir University, School of Foreign Languages
kucuktekeli@gmail.com*

Eray YILMAZ, Ph. D.

*T.C. Ziraat Bankası Balıkesir Fen Lisesi
eray_yilmaz@yahoo.com*

Abstract

The Common European Framework of Reference for Languages (CEFR) is a description of language ability which has been used across Europe and in some other countries of the world. The main purpose of CEFR is to provide a common method of teaching, learning and assessing which applies to languages spoken in Europe. The current study aimed to develop a grammar achievement test in A2 level in accordance with CEFR. The grammar achievement test which was developed in order to find out the validity and reliability was conducted on the 98 students of prep classes in the last week of fall semester in 2014-2015 academic years. At first, an item pool consisted of 64 items was prepared by the researchers. Then, in order to ensure the validity and reliability of the test, the reviews of two field experts were taken and the number of items that the draft test would be consisted of was identified. Afterwards, the analysis of item discrimination and item difficulty analysis were carried out. As a result of the item discrimination analysis 12 items and after the item difficulty analysis 4 items were taken out from the draft test which had consisted of 50 items previously. Thus, the ultimate test was composed of 34 items. In order to find out the internal reliability, KR-20 value of the test was identified and as a result of the analysis it was found that the test was reliable with the value of .90. The results revealed that the test was valid and reliable enough to be conducted as a measuring tool.

Keywords: Grammar achievement test, The Common European Framework of Reference for Languages, item discrimination, item difficulty, validity

The Council of Europe (COE) which was founded in 1949 has currently 47 member states. A series of regulations have been carried out in many areas with the joint collaboration of COE and European Union (EU). Language education is one of the education policies aim to promote plurilingualism, linguistic diversity, mutual understanding, democratic citizenship, and social cohesion (COE, 2006). As Güler (2005) stated, improving mutual understanding, establishing collective consciousness among European citizens who have different cultural backgrounds can only be available with common communication systems. Therefore, CEFR is an important mean to overcome the barriers in communication resulting from different educational backgrounds. It provides a common basis for the elaboration of language syllabuses, curriculum guidelines, examination, textbooks, etc. across Europe (COE, 2001). As Fulcher (2004) stated the CEFR is increasingly being used as a standard document for language teaching and assessment in Europe. CEFR describes what language learners have to learn in order to use a language for communication and what knowledge and skills they have to develop to be able to act and interact effectively (COE, 2001). For Heyworth (2004), language teachers, teacher trainers and academic managers need to consider CEF as a comprehensive reference for language education along with its practical applications.

CEFR also provides a measurable standard that can be associated with 6 descriptive levels of learning which are A1, A2, B1, B2, C1 and C2. Each descriptor is expressed in “can do” statements and follows the order reception, production, interaction. Each level is classified with descriptors of different skills, understanding, speaking and writing. It also provides a rating scale for spoken proficiency, and assessment criteria for range, accuracy, fluency, interaction and coherence in the areas of linguistic, socio-linguistic, strategic and pragmatic competence. As CEFR is based on action oriented and focuses on communicative competence of learners, grammar is not mentioned much in it. For Rama and Agullo (2012), there are number of unsolved issues in CEFR related with the grammatical competence, regarding what grammar is and its role in language learning. However, since each language has complex and unique rules in its own context, CEFR also stated that it is not possible to produce a scale for all languages for progression in respect of grammatical structures. There are several theories for sentence formation. Therefore, the framework does not decide which to use (COE, 2001). North (2004) also stated that CEFR tries to prescribe how it should be taught, rather than what should be taught. In the document grammatical competence was defined as knowledge of, and ability to use, the grammatical resources of a language and the

ability to understand and express meaning by producing and recognizing well-formed phrases and sentences in accordance with these principles (COE, 2001).

From the point of testing and assessment, it would be unfair to say that CEFR has some drawbacks in terms of language testing and assessment since the Council of Europe produced a manual consisted of 188 pages called *Relating Language Examinations to the Common European Framework of Reference for Languages: Learning, Teaching and Assessment (CEFR)* in 2009. Reliability and validity are the vital factors in test quality. In the document, three fundamental concepts of assessment were explained. Validity is the most important quality of test use and meaningful inferences can be drawn from test scores through it (Bachman, 1990). Moreover “reliability concerns the extent to which an experiment, test or any measuring procedure yields the same results on repeated trials” (Carmines & Zeller, 1979, p. 17). As the focus of this research is on validity and reliability of a multiple choice achievement grammar test, the adequacy and the achievements of students for the intended course can be obtained through achievement tests (Brown, 1996). Moreover, the preparation of the items are of great importance as it was stated by Hughes (1989) it is really difficult to write good multiple choice items, much time and effort is needed for the construction. Furthermore, multiple choice tests have been criticized by many scholars for not evaluating the abilities of test takers’ communicative and productive skills and encouraging the test takers on guessing the right item without having required knowledge on the subject. However, multiple choice tests were thought to be the ideal measuring tools to evaluate the students’ knowledge of vocabulary and grammar (Harmer, 2001). Even today they are being used widely in several areas of educational activities. They provide fast and simple assessment and enable the tests to be more reliable than subjectively evaluated tests for being evaluated objectively (Bailey, 1998). Besides, multiple choice tests are a good alternative for increasing reliability (Argüelles et al., 2012).

As North (2004) stated CEFR levels match up well with to the Alte and Cambridge Esol levels. Therefore, A2 level in CEFR is accepted matching with (Key English Test) KET. Although in the document the outcome of the grammar accuracy of A2 has been stated as “uses some simple structures correctly, but still systematically makes basic mistakes – for example tends to mix up tenses and forget to mark agreement; nevertheless, it is usually clear what he/she is trying to say” (COE, 2001, p. 114). As it was mentioned earlier, since CEFR does not state what should be taught, the grammar topics were determined from a corpus of texts and general English course books in accordance with KET and CEFR.

Method

The current study focused on three main research phases; development process of a grammar achievement test, validity and the reliability of the test.

Development of items pool and study of validity

A description table with regard to Bloom's taxonomy of educational objectives was developed as in the similar study carried out by Adıgüzel and Özüdođru (2013). Grammar topics of A2 level were arranged in the description table in accordance with Bloom's cognitive levels. As the achievement test consisted of multiple choice items, there were no items related with synthesis and evaluation levels. Description table consisted of knowledge, comprehension, application and analysis levels were given in Appendix A.

An items pool consisted of 64 items was developed in accordance with Bloom's taxonomy of educational objectives. Two field experts were consulted to ensure the content validity of grammar topics in A2 level. As a result, it was decided to develop the draft test consisted of 50 items, subtracting 14 items out of 64.

Reliability analysis of draft achievement test

For the reliability analysis of draft achievement test, data were collected from the prep class students at the University of Balıkesir in 2014-2015 education periods. Draft achievement test consisted of 50 items was conducted on 98 students who were expected to be A2 level at the last week of fall semester. Through item analysis, item discrimination and item difficulty indices were calculated.

Thanks to item discrimination index (D), individuals on high and low levels can be discriminated concerning the feature that is being measured. In case of item discrimination index which can vary between -1 and +1 is negative, the related items must be removed from the test due to reverse discrimination of individuals (Büyüköztürk et al, 2010). In item discrimination analysis process, the total scores of the participants gathered from the scale were sorted from highest to lowest. T-test is carried out between the highest group of 27% and the lowest group of 27% in order to identify whether the difference is significant (Ergin, 1995).

Following the sorting of the students' scores in the test from highest to lowest, two separate groups each of which composed of 27 students ($98 \times 27 / 100$) were constituted. Item discrimination index was calculated with regard to responds of the participants for each question in the achievement test. SPSS 16.0 was used to perform analysis and the questions which had higher significance level than .05 were taken out from the achievement test.

Item difficulty index (P) is a measure of the proportion of the participants who answer the each item correctly. This proportion is found with the division of participants who answered correctly by all participants and it can range between 0.00 and +1.00 (Baykul 2010). If a greater proportion of participants respond the question correctly, item difficulty index closes to +1 and it is thought to be an easy question. If only a small proportion of participants respond the question correctly, item difficulty index closes to 0 and the question is thought to be difficult (Özçelik 2010). It must be paid attention to have the item difficulty index is between .20 and .80 (Özçelik, 1992).

Item difficulty index (P) was calculated for each item in the achievement test. Questions which had item difficulty values below .20 were determined as very difficult and the ones which had values above .80 were determined as very easy and, they were taken out from the achievement test. Questions were classified in terms of difficulty indices and were evaluated as between $.61 \leq P \leq .80$ was easy, between $.41 \leq P \leq .60$ was moderate and between $.20 \leq P \leq .40$ was difficult. In order to identify the internal reliability, KR-20 value of the test was calculated.

Findings

The findings obtained through the analysis and the interpretations of the findings were mentioned in this section of the study. Independent t-test was carried out in order to identify item discrimination of highest and lowest groups of 27%. The results of the analysis were given in Table 1.

Table 1
Item Discrimination Indices of Questions

	F	Sig.	t	p		F	Sig.	t	p
Q1	18928.000	.000	5.292	.000	Q26	16.863	.000	2.686	.010
Q2	5.046	.029	6.523	.000	Q27	360.816	.000	3.911	.001
Q3	7.255	.009	1.306	.198	Q28	1.163	.286	1.941	.058
Q4	16.811	.000	-1.850	.071	Q29	1.504	.226	3.976	.000
Q5	7.255	.009	4.896	.000	Q30	1.368	.247	3.228	.002
Q6	2.628	.111	.845	.402	Q31	5.489	.023	3.260	.002
Q7	15.253	.000	2.096	.041	Q32	31.115	.000	4.209	.000
Q8	53.520	.000	3.053	.004	Q33	23.122	.000	2.473	.017
Q9	.817	.370	.539	.592	Q34	31.115	.000	4.209	.000
Q10	.356	.553	3.581	.001	Q35	10.998	.002	4.479	.000
Q11	43.059	.000	2.750	.009	Q36	15.231	.000	1.890	.065
Q12	25.226	.000	2.772	.008	Q37	8.679	.005	5.953	.000
Q13	3.127	.083	1.375	.175	Q38	43.059	.000	5.157	.000
Q14	.231	.633	-.536	.594	Q39	3.250	.077	3.606	.001
Q15	19.510	.000	3.738	.000	Q40	6.261	.016	4.019	.000
Q16	2.926	.093	2.896	.006	Q41	32.967	.000	2.451	.019
Q17	36.350	.000	2.891	.006	Q42	1.163	.286	1.941	.058
Q18	732.160	.000	6.150	.000	Q43	360.816	.000	6.648	.000
Q19	2.628	.111	2.252	.029	Q44	.310	.580	2.239	.029
Q20	154.025	.000	4.243	.000	Q45	1.688	.200	.645	.522
Q21	39.406	.000	3.519	.001	Q46	.392	.534	4.377	.000
Q22	154.025	.000	4.243	.000	Q47	732.160	.000	4.228	.000
Q23	16.863	.000	2.387	.021	Q48	31.115	.000	4.209	.000
Q24	10.119	.002	1.749	.086	Q49	1.261	.267	.568	.572
Q25	43.059	.000	2.750	.009	Q50	1.688	.200	4.837	.000

According to t-test results in Table 1, it was found that 10 questions which did not have significant difference between the lowest and highest groups were not discriminative enough. The questions 3, 6, 9, 13, 24, 28, 36, 42, 45 and 49 which had lower discriminative levels were taken out from the achievement test. Since T values of question 4 and 14 were calculated as -1.850 and -.536 respectively they were taken out from the test as well for the reason of reverse discrimination.

After item discrimination analysis 12 questions were taken out from the test and the rest of the test was composed of 38 questions. Item difficulty indices and results of the analysis of the 38 questions were given in Table 2.

Table 2

Item Difficulty Indices of Questions

No	Right	Wrong	P	No	Right	Wrong	P
27	83	15	.85	35	62	36	.63
47	81	17	.83	2	61	37	.62
8	80	18	.82	30	54	44	.55
25	79	19	.81	40	54	44	.55
11	76	22	.78	46	52	46	.53
12	76	22	.78	5	51	47	.52
32	76	22	.78	10	49	49	.50
17	75	23	.77	44	46	52	.47
22	75	23	.77	19	45	53	.46
48	72	26	.73	29	40	58	.41
15	69	29	.70	16	39	59	.40
34	69	29	.70	50	38	60	.39
1	68	30	.69	39	36	62	.37
31	68	30	.69	23	35	63	.36
21	67	31	.68	26	32	66	.33
43	67	31	.68	33	31	67	.32
18	66	32	.67	20	25	73	.26
37	64	34	.65	7	21	77	.21
38	63	35	.64	41	21	77	.21

According to Table 2 there were not many difficult questions in the test ($P \leq .20$). The questions 8, 25, 27 and 47 were identified as very easy questions for having item difficulty indices higher than .80 and they were taken out of achievement test. After the calculation of item difficulty indices 4 more questions were taken out from test and the final test composed of 34 questions. Difficulty distributions of the questions were given in Table 3.

Table 3

Distributions of Questions in Final Achievement Test With Regard To Item Difficulty Levels

Difficulty Value	Number of Questions	Sequences of Questions	Difficulty Level
$.61 \leq p \leq .80$	17	1,2,11,12,15,17,18,21,22,31,32,34,35,37,38,43,48	Easy
$.41 \leq p \leq .60$	8	5,10,19,29,30,40,44,46	Moderate
$.20 \leq p \leq .40$	9	7,16,20,23,26,33,39,41,50	Difficult

According to Table 3, as the questions in the final achievement test were in different difficulty levels, they have the capability of measuring students' achievements in different levels.

In order to identify the internal reliability of achievement test, KR-20 (Kuder-Richardson) value was calculated. Test scores having reliability coefficient .70 or higher are generally considered to have adequate reliability values (Büyüköztürk, 2004). According to result of the analysis, internal reliability of the test was calculated as .90 and it was found that the test is reliable enough to be conducted.

Results and Discussion

This study aimed to develop a valid and reliable grammar achievement test consisted of multiple choice items in A2 level in accordance with CEFR. Item discrimination and item difficulty indices of the test were performed through analysis. As a result of the research, an achievement test having higher validity and reliability values was developed. The achievement test was consisted of 34 questions in easy, moderate and difficult levels.

The achievement test can be used readily for A2 level grammar proficiency of the students as a measuring tool by the English language teachers. Besides, the test can be used as pretest or posttest in experimental studies for academic purposes.

Further studies can be carried out by reaching more participants. Interviews can be conducted with the students on the identifications of the difficulty levels of the questions to obtain qualitative findings for the further studies. As CEFR was adopted by many countries, curriculum was designed in accordance with this framework in schools and universities. Thus, such achievement tests are of great importance as a tool for measuring students' proficiencies.

References

- Adıgüzel, O. C., & Özüdođru, F. (2013). Üniversitelerde ortak zorunlu Yabancı Dil-I dersine yönelik bir akademik başarı testinin geliştirilmesi. *Trakya Üniversitesi Eğitim Fakültesi Dergisi*, 3(2), 1-11.
- Argüelles Á. I., & Lerchundi, I. P. (2012). ... and back to multiple choice! Large-scale testing of proficiency in English: An experience. *Odisea*, ISSN 1578-3820, 9-18. Retrieved December 21, 2014, from <http://hdl.handle.net/10835/2074>
- Bachman, L. (1990). *Fundamental considerations in language testing*. Oxford: Oxford University Press.
- Bailey, K. M. (1998). *Learning about language assessment*. Boston: Heinle & Heinle Publishers.
- Baykul, Y. (2010). *Eđitimde ve psikolojide ölçme: Klasik test teorisi ve uygulaması (2. Baskı)*. Ankara: Pegem Akademi, 11-31.
- Brown, J. D. (1996). *Testing in language programs*. New Jersey: Prentice Hall Regents.
- Büyüköztürk, Ş. (2004). *Veri analizi el kitabı (4.Baskı)*. Ankara: Pegem Akademi Yayıncılık.
- Büyüköztürk, Ş., Çakmak, E. K., Akgün, Ö. E., Karadeniz, Ş. & Demirel, F. (2010). *Bilimsel araştırma yöntemleri*. Ankara: Pegem Yayıncılık.
- Carmines, E. G. & Zeller, R. A. (1979). *Reliability and validity assessment*. Newbury Park, CA: Sage Publications.
- COE. (2006) *Plurilingual education in Europe*: Retrieved October 12, 2015, from http://www.coe.int/t/dg4/linguistic/Source/PlurinlingalEducation_En.pdf

- COE. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Ergin, D. Y. (1995). Ölçeklerde geçerlik ve güvenilirlik. *Marmara Üniversitesi Atatürk Eğitim Fakültesi Eğitim Bilimleri Dergisi*, 7, 125-148.
- Fulcher, G. (2004). Deluded by artifices? The Common European Framework and harmonization. *Language Assessment Quarterly*, 1(4), 253-266.
- Güler, G. (2005). Avrupa Konseyi Ortak Dil Kriterleri Çerçeve Programı ve Türkiye’de yabancı dil öğretim süreçleri. *Trakya University Journal of Social Science*, 6(1), 89-106.
- Harmer, J. (2001). *The practice of English language teaching*. Harlow: Pearson Education.
- Heyworth, F. (2004). *Why the CEF is important?* In K. Morrow (Ed.). *Insights from the Common European Framework* (pp. 12-21). Oxford: Oxford University Press.
- Hughes, A. (1989). *Testing for language teachers*. Cambridge: Cambridge University Press.
- North, B. (2004). Europe’s framework promotes language discussion, not directives. *The Guardian*. Retrieved November 7, 2014, from <http://www.theguardian.com/education/2004/apr/15/tefl6>
- Özçelik, D. A. (1992). *Ölçme ve değerlendirme*. Ankara: ÖSYM.
- Özçelik, D. A. (2010). *Test hazırlama kılavuzu (4.Baskı)*. Ankara: Pegem Akademi, 11-29.
- Rama, J. L. & Agulló, G. L. (2012). The role of grammar teaching: From communicative approaches to the Common European Framework of reference for languages. *Revista de Lingüística y Lenguas Aplicadas*, 177-191.

Appendix A

Description Table of Draft Test with Regard to Bloom's Taxonomy of Educational Objectives

SUBJECT	Knowledge	Comprehension	Application	Analysis	Synthesis	Evaluation
To be Subject Pronouns/Demonstratives/ Singular/Plural Nouns	1		2	1		
Present Simple Adverbs of frequency	1	1		1		
Possessive Adjectives Possessive Pronouns Possessive's			1			
Adverbial Phrases of time and place/ Prepositions	2			1		
Present Continuous Present Continuous for future		1				
Present Simple vs. Present Continuous			1	1		
Object Pronouns Reflexive Pronouns		1		1		
Modals: Can / Can't		1				
Past Simple Was/were	1		1			
Past Simple Regular/ Irregular Verbs	1		1			
Countable and Uncountable Nouns Quantifiers		1	1			
Modals: must/have to/should				3		
Adjectives Comparative/Superlative		2				
Future Be going to/Will				2		
Adverbs		1	1			
Phrasal Verbs Separable/inseparable	2					
Past Continuous When/while		1	1			
Asking for permission Offer Suggestion	1	2	1			
Present Perfect	1	1				
If Clause Zero Conditional	1	1				
Gerunds/Infinitive			2			
Question Words	2	2				
Total	13	15	12	10		

Appendix B

A2 Level Grammar Achievement Test

Dear Students,

This test was developed for scientific research. It contains 34 questions which focus on grammar at A2 Level. Your responses to this test will be kept confidential and will have no effect on your academic score. Duration of the test to answer the questions is 30 minutes.

Choose the Best Answer.

1- A: How much are _____ socks?

B: _____ €5.

A)that/They are B)this/These are C)those/It is D)these/They are E)they/It is

2- Please don't make noise. Your father _____ a nap.

A)take B)is take C)took D)takes E)is taking

3- _____ you study hard, you can't be successful.

A)Unless B)When C)While D)If E)For

4- Teachers work _____ but they don't earn much money.

A)fastly B)hardly C)hard D)difficult E)good

5- A: What is in the newspaper today?

B: I don't know. I _____ it, yet.

A)didn't read B)haven't read C)don't read D)won't read E)am not reading

6- Is that _____ car? – No, _____ is red, not blue.

A)mine/your B)their/they C)yours/my D)they/it E)your/mine

7- Sue always _____ her homework in the library, but today she _____ at home.

A)do/is studying B)does/is studying C)is doing/is studying D)is doing/studies
E)does/studies

8- You _____ talk in the library.

A)don't have to B)mustn't C)have to D)may not E)should

9- There _____ any petrol in the tank.

A)was B)is C)are D)aren't E)isn't

10- His plane is arriving _____ half past eleven _____ Saturday night.

A)at/on B)on/at C)at/in D)on/in E)at/at

11- I'm going shopping because we need _____ eggs and _____ kilo of cheese.

A)some/some B)an/a C)some/a D)any/a E)an/some

- 12- _____ you take an aspirin if you have a headache?
 A) Why aren't B) While C) Should D) Let's E) Why don't
- 13- A: _____ was the weather like yesterday?
 B: It was sunny.
 A) How B) Which C) Where D) When E) What
- 14- There are two _____ and a _____ in the car.
 A) Man/woman B) Women/men C) Women/man D) Men/women
 E) Mans/women
- 15- Where _____ you stay when you were on holiday?
 A) did B) have C) were D) do E) are
- 16- How _____ does it take to go to Venice from Florence?
 A) many B) much C) often D) far E) long
- 17- The exam was difficult but I did _____.
 A) good B) quick C) easy D) best E) well
- 18- You _____ wash that shirt because it isn't dirty.
 A) mustn't B) don't have to C) can't D) should E) may
- 19- I will take _____ my jacket. It is hot in here.
 A) off B) out C) for D) in E) up
- 20- We were having dinner when the phone _____.
 A) rings B) rang C) was ringing D) has rung E) is ringing
- 21- They didn't _____ the match last night.
 A) watched B) watch C) watches D) watching E) to watch
- 22- I'd _____ watch basketball than play it.
 A) better B) more C) rather D) faster E) like
- 23- A: _____ to have some cold water?
 B: No, thank you. I am not thirsty.
 A) Did you want B) Do you like C) Would you like D) Can you E) May you
- 24- _____ does Nancy go to the gym? Twice a week.
 A) When B) How C) How long D) How often E) How far
- 25- A: _____ have you been driving?
 B: For eleven years.
 A) When B) How often C) How long D) How many E) How far
- 26- I saw Lisa's brother in town today, but I don't think _____ saw _____.
 A) she/my B) I/him C) him/his D) he/me E) she/him

- 27- John_____ everybody in the town because he_____ for a long time.
A)has known/was living B) knows/has lived C)knew/lives D)knows/lives
E)know/is living
- 28- Bob_____ solve this problem, because he is good at mathematics.
A)should B)don't have to C) may D) has to E)can
- 29- A: Who_____ your mother in the kitchen?
B: Usually my sister.
A)helps B)does help C) is help D)did help E)do help
- 30- Soldiers_____ wear uniforms.
A)can B)don't have to C)should D)may E)have to
- 31- When people are alone, they often talk_____.
A)them B)their C)themselves D)theirselves E)himself
- 32- I was ill so I_____ at school yesterday.
A)didn't B)don't C)wasn't D)am not E)be
- 33- I talked to Liz on the phone yesterday. She is_____ visit us next Saturday.
A) be B) go to C) will D) going to E)has visited
- 34- She likes_____ tennis, but she doesn't want_____ now.
A)plays/playing B)to play/playing C)plays/played D)playing/to play
E)playing/playing

Key

- | | | | |
|-------|-------|-------|-------|
| 1. D | 11. C | 21. B | 31. C |
| 2. E | 12. E | 22. C | 32. C |
| 3. A | 13. E | 23. C | 33. D |
| 4. C | 14. C | 24. D | 34. D |
| 5. B | 15. A | 25. C | |
| 6. E | 16. E | 26. D | |
| 7. B | 17. E | 27. B | |
| 8. B | 18. B | 28. E | |
| 9. E | 19. A | 29. A | |
| 10. A | 20. B | 30. E | |