

Evaluation of 10th Grade English Textbook Activities Using Rasch Analysis

Murat POLAT

Anadolu University, Türkiye

Abstract

The purpose of this study is to examine the data collected using a scale to be developed in order to determine the functionality of foreign language activities in the 10th grade English textbook used by Turkish Ministry of National Education, using the Multi-Facet Rasch Model (MFRM). The study aims to focus on the extent to which the CEFR (Common European Framework of Reference for Languages) and holistic education approach, which are particularly emphasized in the “Turkish Century Education Model”, are reflected in the content of the language book. The 10 units of the book were evaluated by a nine-person evaluation board consisting of three experts from foreign language education, program development, and assessment & evaluation disciplines. These experts examined the content of the book using the “Textbook Evaluation Rubric” which was developed particularly for this study. Furthermore, it was primarily aimed to determine how effective the foreign language activities in the book would be when the purposes of skills acquisition and holistic education that might bring different skills together were observed, and what they promised in terms of usability and difficulty levels for language learners was analysed. Finally, the functionality of the category levels of the textbook evaluation rubric were checked and whether they were at a statistically acceptable level were tested and identified. The last but not the least, in the light of the results compiled from the study; how effective and explanatory MFRM could be in such empirical studies where there are many surfaces and each surface is intended to be examined in detail have been discussed based on the findings obtained from each research phase.

Keywords: *Education model, English textbook, textbook review, textbook evaluation rubric, MFRM*

INTRODUCTION

Education is one of the most important processes that shapes the future of a society. The Maarif Model, developed in line with Turkey's future vision and Turkey Century goals, aims to raise generations who are committed to national and spiritual values, equipped with universal knowledge and skills, and capable of meeting the needs of the future. In this context, foreign language education and especially English teaching has critical importance, because it is very important for new generations to be able to speak a foreign language well in order to be prepared for the world and the life ahead of them. However, the problems experienced in foreign language education in Turkey for many years have been voiced by all stakeholders for some time, and despite the time and resources spent, the targeted education level has not been achieved not only in English but also in the education of many other foreign languages. The problem is of such a scale that it cannot be solved quickly and with hasty measures. Indeed, it has been observed that the changes and innovations made by both the Ministry of National Education and other institutions within their own structures have not been of much use, and have left negative feedback and negative impressions on almost all stakeholders regarding foreign language learning. Unfortunately, Turkish students have begun to believe that learning English is very difficult, expensive, and even a complex task that only a certain group can accomplish, unless they are included in an intensive language education program, cannot take lessons from good English teachers, cannot use expensive language books, and sometimes cannot find the opportunity to practice abroad (Bektaş & Çetinkaya, 2020; Işık & Işık, 2017). This situation seems unacceptable based on the fact that English is the easiest foreign language to learn, because if all nations of the world can easily learn and speak this language, Turkish students can also learn it and even speak it better than them. Therefore, the problem needs to be addressed.

In this study, the quality of foreign language textbooks taught in schools in Turkey and the extent to which they can respond to today's needs will be examined. For this purpose, it is aimed to determine the impact of foreign language activities in the 10th grade English textbook used by the Ministry of National Education (MEB) on

language teaching. The research will particularly focus on how well the skills framework and holistic education approach (Turkish Maarif Foundation, 2023), which have been emphasized within the "Turkish Century Maarif Model," are reflected in the book. The study has three main objectives. First, an analytical rubric will be developed by obtaining opinions from experts to be used in English textbook review. The second objective of the study is to have experts examine the content of the 10th Grade English textbook and determine how effective the foreign language activities in the said book will be considering skill acquisition and holistic education objectives that will evaluate different skills together. It will be determined to what extent the reading, listening, writing and speaking activities in the ten units of the book will be useful in terms of usability and difficulty levels, considering the foreign language levels of 10th grade students. The third and final objective of the study is to test the functionality of the category levels of language activity criteria included in the rubric to be developed and whether they are at a statistically acceptable level using infit and outfit values. Additionally, in light of the results compiled from the study, the extent to which the Many-Facet Rasch Model (MFRM) can be effective and explanatory in empirical studies where multiple facets and each facet is desired to be examined in detail will be discussed based on the findings obtained, and suggestions will be presented for possible future studies to be conducted with MFRM.

Conceptual Framework

Among the education programs that yield the least efficiency in the education content provided in public schools in Turkey, foreign language education ranks first, and despite the regulations, teacher support and innovations made for years, the desired results cannot be achieved (Alptekin, 2021; Işık & Işık, 2017). For example, while Turkey is in the same category as third world countries in terms of English proficiency, it ranks last in Europe (Polat, 2020). Of course, this problem has different causes, but we can think that starting with the materials and activities used in foreign language lessons would be finding a solution to an important aspect of the problem.

As we remember from our student years, foreign language textbooks are mostly a reference source (Council of Europe, 2020; Kachru, 2005; Stufflebeam, 1972). They can be a good reference for someone with limited means, or both a guide and a reference point for intensive language education programs or novice instructors (Cunningsworth, 1995). From a more global perspective, especially in countries where English is taught as a foreign language, the quality and importance of textbooks that directly affect foreign language education is indisputable, as these resources are the cornerstone of foreign language education programs (Solak, 2011). Therefore, they are not only expensive but also rare and somehow frequently subject to changes (Binbaşıoğlu, 1994). Millions of these materials used in foreign language education are produced and marketed every year, and different versions of the same-named book with changes in their content for different countries are priced according to the economic conditions and foreign language education dynamics of the country where they are sold. No learning or teaching environment is complete without relevant materials, support materials or textbooks. O'Neill (1982) suggested that regarding English books, either a lot of money should be paid for quality ones and the cost of language education should be borne, or institutions should write their own books or create workbooks with their own resources.

The books used in schools in Turkey have been provided free of charge to children and young people by the state for a long time, thus providing significant unity and savings in the use of educational materials (Ceyhan, 2021). The quality of these books used, the richness of their content and their ability to meet needs have gained more importance with this practice, as there is no other option that can be used by teachers in lessons. Therefore, it would be beneficial for all textbooks taught in schools to be examined by independent researchers and to provide feedback on their shortcomings (Akdeniz et al., 2002). For this reason, in this study, the content of the 10th grade English book used as a textbook in schools, how efficient the speaking, writing, listening and reading activities can be, and to what extent they can contribute to the foreign language education of Turkish students will be discussed. Based on the expert opinions to be obtained as a result of the study, the units in the book will be addressed with a holistic evaluation, and their strengths and weaknesses will be revealed.

LITERATURE REVIEW

When the literature on the subject is examined, it emerges that the foreign language books intended to be taught in lessons should be presented in a way that combines some qualities emphasized in the Turkey Century Maarif Model (Turkey Maarif Foundation, 2023) and the knowledge and skills desired to be imparted to Turkish students with culture, customs and traditions, along with the spiral education and communication-oriented language skills required by the age (Akdoğan, 2022; Ceyhan, 2021; Demirel, 2020, Solak, 2011). In this context, the expected characteristics of a good English textbook are:

1. Balances national and universal values,
2. Parallel to the integrated development of language skills,
3. Oriented towards technology integration and digital competencies,
4. In a gradual and spiral structure,
5. Open to values education and character development,
6. Provides diversity in measurement and evaluation,
7. Suitable for interdisciplinary approach,
8. Has intercultural awareness and global perspective,
9. Student-centred and individualized content,
10. Should have content compatible with Turkey's strategic goals MEB (2023a).

In today's world where internet use is an inseparable part of daily life and communication can be established with even the most distant countries within seconds, it is very important for young people to preserve their cultural identities despite intense external influences, to become familiar with texts and content reflecting Turkish history, culture and values; to read about the achievements of Turkish scientists, artists and historical personalities from different sources but in a foreign language, and to listen to content introducing Turkey's geographical, cultural and touristic riches from, for example, an English-speaking guide (Demirel, 2020). When the fact that learning a language is also learning the world is remembered, it is an indispensable requirement of today's age for young people to gain universal perspectives, develop global citizenship consciousness, and receive education that develops respect and understanding for different cultures (Turkey Maarif Foundation, 2023). The basic education model should be the education of individuals who develop intercultural communication skills on one hand, and on the other hand are aware of important global problems and can use the target language at a level where they can present their own solution suggestions and discuss them comfortably (MEB, 2023a; YÖK, 2023). A good foreign language book should include:

- Listening materials taken from real life, appropriate to the level and including different accents,
- Functional dialogues that can be used in daily life and activities that provide speaking practice,
- Reading passages that contain interesting cultural elements and encourage critical thinking,
- Gradual activities that develop writing skills,
- Activities that encourage critical and creative thinking and are problem-solving oriented (Allwright, 1982; Kırkgöz, 2009).

In addition to the features mentioned above, Cunningsworth (1984) and Göçer (2021) suggested that good language education should be designed in a spiral structure at levels that address students' ages and abilities. For example, while content starting with game-based, visually-oriented, concrete concepts should be presented at primary school level, topics related to daily life and reflecting students' interests should be included at secondary school level, and activities that develop academic language skills with more reading and writing content and are profession and career-oriented should be included at high school level. For the permanence of language structures learned in lessons, it is recommended that these be repeated in new contexts, progress gradually from difficult to easy, and have activities that reinforce previous knowledge in each unit and enable the use of language structures in different contexts (Demirel, 2020; Genç, 2023, İşisağ, 2020).

If today's language education also employs valid and reliable language exams where 4 skills are effectively measured, foreign language books should contain the test techniques of such measurement methods (Brown & Abeywickrama, 2019; Karaman, 2021). Moreover, the materials used in language lessons are expected to be suitable for project-based evaluations, skill monitoring scales compatible with the European Language Portfolio, and self-assessment, peer assessment, and portfolio writing, in addition to traditional exams to measure students' development from different aspects (Demirel & Diker-Coşkun, 2021; Fulcher & Davidson 2019). Furthermore, attention should be paid to the quality of books to provide feedback in addition to measurement. To more rigorously monitor and support the development of language learners, books should provide self-assessment sections at the end of each unit, checklists that enable seeing strengths and areas open to improvement, and assessment tools that provide constructive rather than critical feedback (Nakamura, 2002; Sarıçoban & Öztürk, 2024).

Another important expectation that is felt to be lacking in most textbooks but is not much voiced in Turkey is the CLIL (Content and Language Integrated Learning) approach. Ellis (2019) mentioned the importance of books that integrate foreign language learning with content learning of different disciplines such as science, mathematics and social sciences, and Tomlinson (2003) emphasized that books that combine their subjects with English teaching add significant power to foreign language education when fed with English instructions for experiments and observations. Some researchers have also advised that English materials should include basic terminology specific

to different professional groups, especially at high school level, and English writing skills that can be used in job applications, in terms of books contributing to other disciplines (Coyle et al., 2010; Demirel, 2020; Şahin & Yıldırım, 2022). In addition to all these requirements, Mirici stated that the performance shown by Turkish youth in foreign language activities based especially on production is not sufficient, that they have problems especially in speaking and writing in the target language (2003), and conveyed that the books to be used in foreign language lessons should include activities of repetition, performance display, role-playing, and task completion, especially in speaking and writing skills. In addition to the content of books used in language lessons, they also need to have many other features such as attractiveness, various visuals, ease of use, price and easy availability (Ersöz, 1990; Cunningsworth, 1995; Solak, 2011).

On the other hand, the inclusion of classic textbooks in English education has been criticized by some researchers and even found unnecessary (Polat, 2020). For example, by drawing attention to the complex relationship between the content of some books used and British culture, it has been argued that these foreign language education books supported by Great Britain prioritize Britain and its ideological connections (Phillipson, 1992). Some researchers believe that these books are sometimes inflexible and impose the educational, psychological and linguistic preferences and prejudices of the authors (Allwright, 1982). For this reason, it can be thought that these books used in language education organize and impose the purpose, method, stages and order of importance in learning the target language (Solak, 2011).

Considering the critical value and role of the textbook intended to be used especially in English education in Turkey mentioned above, there is a need for validity and effectiveness studies to be conducted by the ministry or other relevant institutions, and it should be researched to what extent the currently used versions of these books meet the need, to what extent they keep up with the holistic education understanding we are especially looking for in English education, and how much they comply with the skill-based rather than knowledge-based education understanding explained within the scope of the Turkey Century Maarif Model. Unfortunately, there is not much research in Turkey where foreign language books used by MEB are examined by experts. Şahin and Yıldırım (2022) researched the effect of examining language materials on quality, but they did not focus on specifically selected materials and worked with a general approach. Erdem and Yılmaz (2023), on the other hand, evaluated the communicative approach aspects of MEB English textbooks in their study and did not look at other features. Therefore, this three-stage study, especially by focusing on the 10th grade English book, can contribute to both students and teachers using the book and the authors and editors of the book. In the study, preparing a rubric to evaluate this book, having experts examine the content of the book, and whether the functionality of the category levels of the English activity criteria in the said book and whether they are at a statistically acceptable level will be tested using infit and outfit values. The study will be a pioneer in addressing an important deficiency in the relevant literature and can serve as a reference for academics and teachers who will conduct similar studies in the future.

METHODOLOGY

Mixed method was used in this study. Tashakkori and Creswell (2008) define mixed method as a research in which the researcher collects and analyses data, integrates findings, and makes inferences using both qualitative and quantitative approaches or methods. While a quantitative study is useful for summarizing large amounts of data and reaching statistics-based generalizations, the qualitative method approach is useful for collecting in-depth information for a broad understanding of evaluation and improvement suggestions (Haryanti, 2013). Using only quantitative methods does not guarantee in-depth data, therefore qualitative method was also used to further evaluate the book and provide better and evidence-based suggestions, thus mixed method was preferred (Coombe et al., 2020). In the first stage, ethics committee approval was obtained from the relevant institution, then a "Book Evaluation Rubric" was prepared to collect the expert opinions necessary for the study, and after reliability analyses were conducted, the content of the textbook was examined by experts and the results obtained were analysed using MFRM.

Participants

The study group consists of a total of 20 volunteer participants: 11 English teachers interviewed for content creation for the rubric, and a nine-person-rater group consisting of two measurement and evaluation, two curriculum development, and five English education experts whose opinions were consulted in book evaluation.

Data Collection

In the data collection phase of the research, first, the English activities to be examined by experts were determined. For this purpose, interviews lasting approximately 1 month were conducted with a total of 11 English teachers working in different high schools teaching 10th grade English, with teaching experience ranging from 10 to 22 years. The teachers were asked to select 10 lesson activities they wanted from the textbook that was the subject of the study, and based on the collected preferences, 10 English education activities selected from 10 different units in the book and consisting of the most preferred activities were determined and examined through speaking, listening, writing and reading skills. Additionally, in order to test the Maarif Model content, 4 separate items in the rubric examined the extent to which the Maarif Model requirements developed in line with the Turkey Century goals could be met. After determining which activities would be studied, a set of evaluation criteria were established regarding the qualities sought in them. The criteria were prepared in 5 weeks with the help of one curriculum development, one measurement and evaluation, and two English education specialists. As a result of asking stakeholder opinions for 3 months on the subject, conducting detailed literature review and compiling expert opinions, an "English Textbook Evaluation Rubric" consisting of a total of 17 items was created. The created rubric was examined by two other English educators to be examined in terms of content, grammar and writing quality. With the revisions made as a result of expert opinions, the final version of the "English Textbook Evaluation Rubric" consisting of 15 items was given. The rubric was organized so that experts would give answers to each question in the form of "no – partially – yes", and in order to provide more data, an open-ended section was added at the end of each question where participants could write their ideas, praise or criticism they wanted to add. The prepared rubric was reproduced and given to nine participants so that they would use one in each unit. In the final stage, ten activities selected from ten units in the 10th grade English book were presented for expert opinion.

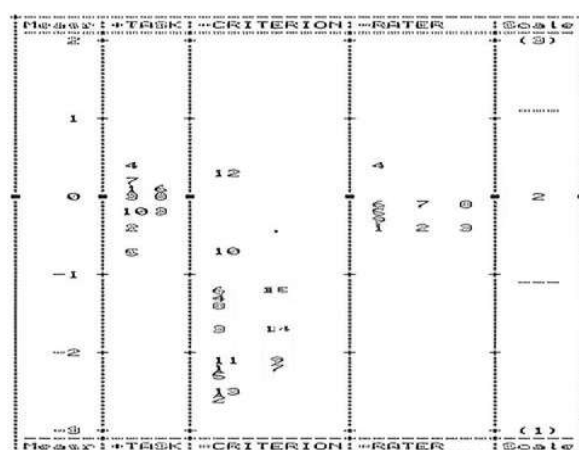
Data Analysis

The descriptive statistics of the 15 items in the prepared rubric were examined and the lowest score obtainable from the rubric was calculated as 0, and the highest score as 30 (no 0, partially 1, yes 2 points). It was observed that the range value for the maximum score obtainable from the rubric was 30. It was determined that the mean of the rubric scores was 17.87, the median was 17, and the standard deviation was 4.03. The skewness coefficient of the score distribution was calculated as -0.411 and the kurtosis coefficient as -0.047, and according to the obtained data, it was seen that the distribution could be evaluated as a normal distribution (İlhan, 2016).

On the other hand, the data obtained from the activities determined using the rubric were analysed using the Many-Facet Rasch Model (MFRM). With MFRM, unlike Classical Test Theory (CTT), instead of evaluating a single variable at once, multiple variables such as rubric item, rater, activity, etc. could be examined simultaneously and independently from each other (Arsan, 2012; Mulqueen et al., 2000; Sudweeks et al., 2005). Thanks to MFRM, it was possible to calculate the compliance of the activities in the English book with the criteria, the consistency exhibited by experts among themselves during evaluation, and the reliability by looking at the distribution of evaluation criteria (Polat et al., 2022). All MFRM calculations were performed using the free version of the FACETS program developed by Linacre (2007).

FINDINGS

The 10th grade English book was evaluated by nine experts on 15 criteria, and the logit scale that emerged as a result of the many-facet Rasch analysis is given in Figure 1. Thanks to MFRM, comparisons regarding the rater, activity and criterion facets considered can be seen on the same linear scale. The "0" point in the middle of the leftmost column of the scale shows the middle of the scores in expert evaluation, the activity corresponding to "0" logit score represents the activity at the middle level according to the criterion in activity scores, and the part showing the rater severity-leniency level corresponding to "0" logit score corresponds to the strictest, middle and most lenient rater in scoring from top to bottom (Semerci, 2011). Thus, someone examining the logit scale has the chance to make comparisons on a linear plane from the English activity that received the highest score to the lowest; from the participant who behaved most strictly while scoring to the most generous; from the criterion with the lowest probability of high scoring for the book as a whole to the one with the highest probability in each column in the figure from top to bottom.

Figure 1: *Wright Map of the 10th Grade English Book Activities*

When the rankings in the logit scale given in Figure 1 are examined, it was seen that the highest scored activity was the activity in the fourth unit (Traditions, Religious Holidays), and the lowest scored activity was in the fifth unit (Travel, 10-Day Anatolia Tour). The third column in Figure 1 contains the averages of the scores given by participants regarding the selected activities of the English book. Accordingly, considering the function of the activity, the item with the lowest probability of high scoring was the 12th criterion (Contains texts that describe Turkey's regional and global role.), while the criterion with the highest probability of good scoring was found to be the second criterion (The words in the textbook are appropriate to the students' level.). Among the participants, according to the scores they gave, raters four gave the lowest scoring, while raters one, two and three respectively gave the highest scoring. The reports of the measurements participating in the analysis regarding possible variables in the evaluation of the book content are presented below in order and in detail. First, the compatibility between the rater and the model and the participants' infit and outfit analysis results were examined.

Table 1: Rater Model Fit Table

Rater	Measure	SE	Infit	Outfit
6 (Assessment & Eval.)	0.78	0.19	0.88	0.91
9 (English)	0.57	0.18	0.68	0.79
7 (English)	0.15	0.18	0.91	0.94
8 (Assessment & Eval.)	-0.09	0.21	0.94	1.03
3 (English)	-0.11	0.22	0.90	0.95
5 (Program dev.)	-0.25	0.22	1.07	0.91
4 (English)	-0.28	0.20	0.86	1.01
1 (English)	-0.33	0.20	0.89	1.03
2 (Program dev.)	-0.45	0.20	0.82	1.13
Mean	0.01	0.18	0.88	0.97
Standard dev.	0.53	0.03	0.18	0.21
Reliability = 0.88 A. Index = 2.53 Chi square= 83.6 df= 8 p = 0.00 Rater fit: 64,6 %				

While examining the measurement values related to the raters' scores presented in Table 1, it can be observed that the sixth rater (participant assessment and evaluation expert) with a logit value of 0.78 was the most stringent scorer compared to other participants, while the second rater (participant curriculum development expert) with a logit value of -0.45 provided the most lenient scoring compared to other participants. Besides the raters' stringency, the infit and outfit data provided in Table 1 show the degree of overlap between the obtained values and the model fit. The outfit statistic is the mean square of residuals between observed data and expected values, and it is a statistic that does not only look at the mean but is quite sensitive to unexpected extreme values (Randal and Engelhard, 2009). On the other hand, the infit statistic is less sensitive to observed extreme values compared to the outfit statistic. The expected value for fit statistics is 1, and this level indicates "perfect fit." Fit statistic values

higher than 1 indicate that the variance of data obtained from raters is greater than expected; values lower than 1 indicate that the variance is less than expected (Fulcher and Davidson, 2019).

If we examine the literature, it can be seen that acceptable values for fit statistics may range between 0.6 and 1.4 according to Lunz et al. (1990), 0.5-1.5 according to Turner (2003), and 0.5-2 according to Myford and Wolfe (2003). In light of these data, it can be said that the data obtained from raters show good fit with the model (Wright and Linacre, 1994; Nakamura, 2002). In summary, lenient measurement values during scoring are observed in raters six (0.78) and nine (0.57). The expertise areas of these raters are assessment and evaluation and English education. The lowest measurement values are observed in raters two (-0.45) and one (-0.33). Rater two's expertise is curriculum development, while rater one's expertise is English.

The standard error values obtained from the analysis are quite close, ranging from 0.18-0.22. While infit values are generally at a good level (ranging from 0.68-1.07), outfit values range from 0.79-1.13. The highest inconsistency among raters is observed in rater two. According to the summary statistics provided at the bottom of Table 1: The reliability coefficient is 0.88 (quite high), the separation index is 2.53, the chi-square value is 83.6, degrees of freedom is 8, and the p-value is 0.00, meaning the fit is statistically significant. Finally, rater agreement is calculated as 64.6% (good level).

While the model generally shows a good reliability level, there are also significant differences among raters, particularly in terms of measurement values. The fact that raters in English and curriculum development fields mostly have negative measurement values indicates that there may be different evaluation approaches in these fields, suggesting they may have different criteria than assessment and evaluation specialists. The Rasch analysis report regarding book activities, another variable included in the study, is presented in Table 2.

Table 2: English Book Activity Measurement Report

Activity	Measure	SE	Infit	Outfit
4	0.47	0.21	1.10	1.24
7	0.24	0.20	1.26	1.22
6	0.15	0.20	1.25	1.17
10	0.10	0.20	1.20	1.11
9	0.09	0.19	1.08	1.01
1	-0.04	0.19	0.90	0.79
3	-0.07	0.19	0.97	0.77
8	-0.09	0.18	1.02	1.12
5	-0.38	0.18	0.98	0.93
2	-0.44	0.17	0.79	0.81
Mean	0.04	0.19	1.03	1.01
Standard dev.	0.28	0.02	0.21	0.24
Reliability = 0.72	A. Index= 1.19	Chi Square = 17.6	df= 9	p = 0.01

The data presented in Table 2 is a report showing the measurement results of English book activities. Based on the data in the table, it can be seen that the activities are ranked according to their difficulty level. The activities found most difficult by the raters are numbers four, seven, and six (with measurement values of 0.47, 0.24, and 0.15, respectively). The easiest activities are numbers two and five (with measurement values of -0.44 and -0.38, respectively). Additionally, standard error values range between 0.17 and 0.21, indicating that the measurements have consistent precision (Wright & Linacre, 1994). When considering infit and outfit values, results are mostly around one or slightly above. Since the ideal value is one, it can be said that there is generally good fit. The reliability coefficient is 0.72, which indicates moderate reliability, and since it is above 0.70, it can be considered reliable (Nakamura, 2002; Randall & Engelhard 2009). The separation index is calculated as 1.19, which shows that the activities have the ability to distinguish different difficulty levels, while the chi-square value is 17.6 and the p-value is 0.01 ($p < 0.05$), indicating a statistically significant result and that the model fits the data well. The mean measurement value is 0.04, which shows that the overall difficulty level of the activities is at a moderate level. In conclusion, this report presented in Table 2 provides an analysis measuring the difficulty levels, compatibility, and reliability of English book activities. The difficulty levels of the activities showed variation and statistically significant results were obtained, with the reliability level also being moderate.

Table 3: Rubric Items Measurement Report

Item	Measure	SE	Infit	Outfit
12	0.43	0.18	0.89	0.88
10	-0.64	0.19	0.91	0.94
6	-1.16	0.20	0.94	0.95
15	-1.16	0.20	0.94	0.96
4	-1.34	0.21	0.79	0.78
8	-1.42	0.21	0.94	0.91
3	-1.71	0.23	1.22	1.24
14	-1.71	0.23	1.22	1.24
9	-2.11	0.26	0.93	0.86
11	-2.11	0.26	1.03	0.97
1	-2.24	0.27	0.95	0.87
7	-2.24	0.27	1.28	1.13
5	-2.29	0.28	1.08	1.02
13	-2.47	0.31	1.39	1.22
2	-2.54	0.31	1.33	0.98
Mean	-2.46	0.24	1.06	1.00
Standard dev.	0.84	0.05	0.20	0.24
Reliability = 0.92	A. Index = 3.21	Chi square = 189.0	df= 14	p = 0.00

The "Rubric Items Measurement Report" presented in Table 3 provides psychometric data for the data collection instrument developed for the research. First, it can be seen that the items in the rubric have difficulty values ranging from -2.54 to 0.43, with the most difficult items being item 12 (0.43, contains texts describing Turkey's regional and global role) and item 10 (-0.64, has content reflecting Turkish family structure and values), while the easiest items are the second (-2.54, The vocabulary in the textbook is appropriate for students' level) and 13th item (-2.54, The variety of topics in the textbook is sufficient). The negative values seen in the measurement section of the items indicate more difficult items. The standard error column shows the measurement precision of each item, and as can be seen, the standard error of the items in the rubric ranges from 0.18 to 0.31. Low standard error means more reliable measurement, and the closer it is to 0, the more precise data the item has provided. Infit values are mostly within acceptable limits (the range of 0.5-1.5 is generally accepted) (Akin & Baştürk, 2012; Fulcher & Davidson, 2019). Only slight misfit is observed in items 3, 7, 13, and 14, which are also within acceptable limits. The general psychometric indicators are: Reliability = 0.92 (excellent internal consistency), Separation Index = 3.21 (good discrimination), Chi-square = 189.0, df = 14, p = 0.00 (statistically significant). While the average difficulty of the items in the rubric was calculated as -2.46 (quite difficult), it exhibited a near-normal distribution with a standard deviation of 0.84 (reasonable spread). All these results indicate that the rubric is psychometrically strong but generally consists of challenging items. The high reliability (0.92) and significant chi-square value confirm that the rubric is consistent and discriminating.

DISCUSSIONS

This study, conducted using a mixed-methods design, carried out a detailed evaluation of the "High School English 10th Grade Student Book" written by Lamia Karamil and published by Pacific Publications according to the Maarif Model. The book aims to develop the English language skills of 10th grade students (ages 15-16, English language lower-intermediate level). Based on the scores given by participants regarding the book and their opinions written in the "Add note" section, results for each unit of the book, inferences made considering the Maarif Century Model, and a summary of the strengths and weaknesses of that unit determined by participants' scores are provided in this section, reflecting the book content along with participants' opinions given with participant codes.

Among the prominent strengths of the book's first unit (School life) is its presentation of realistic dialogues and scenarios related to school life. "Such activities are a good idea for breaking the ice, especially in the first weeks when the course begins, for students to participate in the lesson, speak, and engage in activities" (Participant (P) - 3). Additionally, the activities containing rich vocabulary related to school activities and academic life, comparative presentation of school systems in different countries, and activities that allow students to share their

own school experiences are quite remarkable elements for spiral learning. Ellis (2019) approached this from a similar perspective and mentioned that daily conversation patterns related to school life and implementing this through group work aimed at developing in-class communication skills is an important advantage. Looking at the aspects of the first unit that need improvement, we can first say that the diversity of activities for different learning styles is limited. The opinion that "Activity diversity could have been kept richer at the beginning of the term" was reflected. On the other hand, as Gümüş (2023) also noted, emotional difficulties such as academic achievement and exam anxiety seen in students in Turkey could be addressed, and it could be examined that these are actually problems that may be experienced in every country (P-6). Another issue not given enough importance in this unit is current educational trends such as distance education and digital learning tools. These activities, which cover nearly 1/3 of students' daily lives, could also be a good discussion platform. Finally, the ninth participant conveyed that the difficulty level of reading passages might be above the class level, especially that the text titled "How Do Most American High School Students Spend Their Time?" might be heavy for children in the first week.

Among the prominent strengths of the book's second unit (Plans) is its presentation of meaningful and realistic contexts for teaching future tense patterns. The second participant (curriculum development expert) reported that such activities could enable students to brainstorm about career planning and profession selection because they contain short and long-term goal-setting strategies, and could even lead them to acquire comprehensive vocabulary related to their future plans while doing so. Therefore, among the prominent positive aspects of the second unit are its handling of different types of plans (daily, weekly, academic, career), its interactive inclusion of grammatical structures used to express plans, and its offering of creative writing activities for students to create their own plans. However, insufficient attention to difficulties that may be encountered in the planning process and alternative solutions, and not giving enough weight to profession groups that are particularly gaining importance today, beyond the usual career options presented, were reported as prominent weaknesses. While the fourth participant emphasized that cultural differences' effects on planning habits were not sufficiently addressed, another participant mentioned that there could be more self-evaluation and reflection activities regarding how problematic a generation we have in terms of planning.

When listing the prominent strengths of the third unit (Legendary figure), creating intercultural awareness by focusing on legendary characters from different cultures such as King Arthur, Merlin, Nene Hatun, and Seyit Onbaşı through written and oral English activities, and presenting historical figures with visual materials and interesting narratives were found successful considering the goals of the Maarif Model. On the other hand, in terms of acquiring foreign language skills, the unit content was found successful for effectively processing past tense structures in an effective context, including activities that develop writing and speaking skills about important people and events, and thus presenting discussion questions that develop critical thinking skills that question today's life. However, it was reported that insufficient explanation of the cultural context of some legendary stories would create problems (P-2). A similar finding came from Kiely (2009), where not establishing connections between historical figures and modern heroes and not addressing the reflections of some personalities who have become part of society in today's world was criticized in a source in his study. Another participant (P-7) also reported that the lower number of legendary female figures compared to males could be a sexist choice.

The book's fourth unit (Traditions) was appreciated for its comparative handling of various traditions, cultural values, and ways of expressing them, its emphasis on the importance of belonging, and its provision of content aimed at developing respect and understanding for cultural differences. A similar suggestion came from Byram (2008), who argued that the idea of respect for different cultures should also be developed while teaching English. It was appreciated for presenting rich vocabulary related to cultural celebrations and ceremonies and related grammatical structures in different contexts (P-2,7). However, participants reported that the impact of modern life on traditions could be addressed and that generalizations that risk reinforcing cultural stereotypes in the unit could be problematic (P-4,5). Another view is that traditions are not static social rituals, that they can change, and that speaking and writing activities could be included to prepare students for this evolution (P-8).

The fifth unit (Travel) surpassed other units in terms of teaching practical language use related to learners' future travel needs (making reservations, asking for directions, etc.), providing comprehensive information about different types of travel and destinations, and giving instructions related to travel planning. Polat (2018) in his study praised that an English material should present scenarios aimed at developing intercultural communication skills and contains rich and interesting destination promotions. The first participant also mentioned the concepts of responsible and sustainable tourism included in the unit, emphasizing the importance of including these among foreign language activities. Among the aspects of the unit content that could be characterized as negative are that giving more weight to luxury travel options compared to budget-friendly alternatives creates a handicap and inadvertently directs young people toward such types of travel, and it was emphasized that information about

cultural norms, behavioural rules, and customs in foreign countries that Turks travel to frequently (Germany, Egypt, Russia, etc.) could be provided, thus killing two birds with one stone (P-4,8).

The sixth unit (Helpful tips) was found useful in terms of providing students with practical suggestions they will learn and remember (health, education, technology use, etc.) and containing scenarios for solving small problems on their own and making decisions (P-5). Daloğlu and İnözü (2021) also stated that they found the unit content useful in terms of effectively using advice-giving and receiving structures, students sharing their own experiences, and developing the ability to distinguish appropriate advice for different age and cultural groups. On the other hand, lack of awareness that some suggestions may not be valid everywhere in cultural and individual contexts, the issue of basing such advice on scientific foundations, and the need to focus more on skills for evaluating the reliability of internet-sourced advice were emphasized (P-3,6).

The seventh unit (Food and festivals) was appreciated for its activities, providing content for media literacy skills, different types of news and news analysis techniques, giving English indirect speech structures in a meaningful context, and presenting news-specific vocabulary with rich examples (P-3,4,8). However, it was stated that more could be addressed about the effects of social media, understanding the difference between local and global news sources, and deeper examination of the cultural and political contexts of news was needed (P-1,5).

The eighth unit (Digital era) was appreciated for containing the effects of technology on daily life, practical applications for using and developing digital platforms, presenting current vocabulary related to technology, and introducing technology applications in different fields. On the other hand, it was conveyed that more could be addressed about inequalities in access to technology, technology addiction and negative effects of excessive use, the effects of developments such as artificial intelligence and automation could be discussed more deeply, digital privacy and data security issues could be handled more comprehensively, in short, there could be more discussions about the ethical dimensions of technological developments, and thus the speaking and writing activities in the unit could be made more meaningful (P-2,7,9).

The ninth unit (Modern heroes) was seen to present successful role models in different fields (science, art, sports, etc.), encourage motivation to contribute to society and humanity, encourage students through extraordinary achievements of ordinary people, and motivate students about self-confidence, which they feel lacking today, and conditional structures were thought to be used effectively in this context. However, the unit content was criticized for the need for a more balanced representation rate of women and heroes from different cultures, not addressing the complex and controversial aspects of heroes, and the possibility of deepening the analysis of the media's role in creating hero perception (P-3,8,9).

Finally, the 10th unit (Shopping) was found sufficient with its activities for shopping, asking prices, looking at alternatives, and communicating with shopkeepers, which hold an important place among students' foreign language use purposes, both for asking questions and answering questions with appropriate structures. More writing and speaking activities for today's young people's shopping preferences, where internet shopping and online correspondence are more common, could have been expected. Another criticism is that the words in the unit are generally products suitable for adults' shopping habits (tie, high heels, sleeveless dress, etc.) and do not appeal much to 15-16-year olds (P-5,6).

CONCLUSION

In this study, a detailed evaluation of the 10th Grade English textbook according to the Maarif Model was carried out and its suitability to students' educational needs was examined. In the Turkey Century movement, students are expected not to view English lessons only through the lens of target content like in all other subjects, but while taking this course, they should both learn English and demonstrate skills appropriate to today's educational needs using the target language. As a result of evaluating the said book, it was concluded that its content is aimed at developing the four basic language skills (listening, speaking, reading, writing) in a balanced way, adequately addresses current cultural and social themes that will attract students' interest as desired by the Maarif Model, provides elements that can strengthen intercultural awareness and global citizenship consciousness, presents target grammatical structures in meaningful contexts and communicative approaches in terms of language teaching, has a visually rich and attractive design and content, contains examples reflecting important structures related to current English usage, mostly puts the student at the centre and supports interactive learning.

The aspects of the book that are open to improvement can be listed as more diversified activities for different learning styles and levels, giving more place to critical thinking and problem-solving skills, more balanced

representation of cultural diversity especially in subjects that students focus on, making self-evaluation and reflection activities more numerous and useful, having more flexible content that can be adapted to gradual and spiral education, and finally providing richer content related to subjects and problems that students may encounter in real life. This examination aims to evaluate the book's compliance with the Maarif Model developed in line with Turkey Century goals and to identify its good and improvable aspects. While the book generally appears compatible with the model's student-centered, communicative, and intercultural awareness-developing approaches, it emerged that it could be improved in terms of containing daily life, real language needs, and more oral and written communicative activities.

Recommendations

This study was conducted using a mixed-methods design, and a detailed evaluation of the 10th grade English book according to the Maarif Model was carried out. A total of 20 voluntary participants contributed to the study and produced meticulous, detailed, and good work with their valuable time and opinions. If we were to make some recommendations to other researchers interested in the subject, being generous with time for book review would be the first advice. Since participants have their own workloads, course and exam obligations, at least 6 months should be foreseen for evaluating a book with 10 units. If sufficient time and support are not provided to participants in this regard, the quality of the resulting work will decrease. Another recommendation is to ensure the participation of native English-speaking instructors who also know Turkish in such studies, since it should be known that preparing an English textbook has many dimensions to consider and handling all these dimensions is difficult only by Turkish authors. It can be said that preparing some target structures entirely considering the Turkish education model and supporting them with examples will not be efficient if a native speaker does not help.

This study has been supported by Anadolu University Scientific Research Projects Coordination Unit under the grant number 2659.

Ethics Statement: It is declared that this study complied with the rules outlined in the Higher Education Institutions Scientific Research and Publication Ethics Directive and that no actions were taken that violated scientific research and publication ethics.

Ethics Committee Approval: Anadolu University Social and Human Sciences Scientific Research and Publication Ethics Board' Approval. Document No: 812154

REFERENCES

- Akdoğan, S. (2022). *Yabancı Dil Eğitiminde Materyal Tasarımı*. Ankara: Pegem Akademi.
- Akın, Ö. & Baştürk, R. (2012). Keman eğitiminde temel becerilerin Rasch ölçme modeli ile değerlendirilmesi. *Pamukkale Üniversitesi Eğitim Fakültesi Dergisi*, 31(1), 175- 187.
- Allwright, R., (1982). What do we want teaching materials for? *ELT Journal*, 36/1.
- Alptekin, C. (2021). ELF in ELT: Implications for teacher education and development. *ELT Journal*, 75(2), 197-206.
- Arsan, N. (2012). *Buz pateni hakem değerlendirmelerinin genellenebilirlik kuramı ve Rasch modeli ile incelenmesi*, Yayınlanmamış Doktora Tezi, Hacettepe Üniversitesi: Ankara.
- Avrupa Konseyi (2020). *Diller İçin Avrupa Ortak Başvuru Metni (CEFR) 2020 Companion Volume*. Strasbourg: Council of Europe Publishing.

- Baştürk, R. (2008). Applying the many facet Rasch model to evaluate PowerPoint presentation performance in higher education. *Assessment and Evaluation in Higher Education*, 33(4), 431-444.
- Bektaş-Çetinkaya, Y. (2022). *Kültürlerarası İletişimsel Dil Eğitimi*. İstanbul: Litera Yayıncılık.
- Brown, H. D., & Abeywickrama, P. (2019). *Language Assessment: Principles and Classroom Practices* (3rd Edition). New York: Pearson Education.
- Byram, M. (2008). *From Foreign Language Education to Education for Intercultural Citizenship*. Clevedon: Multilingual Matters.
- Ceyhan, M. A. (2021). Yabancı dil öğretiminde değerler eğitimi. *Değerler Eğitimi Dergisi*, 19(41).
- Coombe, C., Folse, K., & Hubley, N. (2020). *A Practical Guide to Assessing English Language Learners*. Ann Arbor: University of Michigan Press.
- Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Cunningsworth, A. (1984). *Evaluating and Selecting EFL Teaching Materials*. London: Heinemann.
- Daloğlu, A., & İnöz, J. (2021). İngilizce öğretmeni yetiştirmede program geliştirme. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 36(2), 365-378.
- Demirel, Ö. (2020). *Yabancı Dil Öğretimi: Dil Pasaportu, Dil Biyografisi, Dil Dosyası* (12. Baskı). Ankara: Pegem Akademi.
- Demirel, M., & Diker-Coşkun, Y. (2021). Ölçme değerlendirmede alternatif yaklaşımlar. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 36(3), 254-269.
- Demirel, Ö. & Kiroğlu, K. Ed. (2006). Konu alanı ders kitabı incelemesi, (2. Baskı). Ankara: Pegem.
- Erdem, D. ve Yılmaz, F. (2023). MEB İngilizce Ders Kitaplarının İletişimsel Yaklaşım Açısından Değerlendirilmesi. *Eğitim ve Bilim Dergisi*, 48(3), 127-149.
- Ellis, R. (2019). Principles of Instructed Second Language Acquisition. *System*, 33(2), 209-224.
- Fulcher, G., & Davidson, F. (2019). *The Routledge Handbook of Language Testing*. London: Routledge.
- Genç, A. (2023). *Türkiye'de Yabancı Dil Eğitimi ve Planlaması*. İstanbul: Nobel Akademik Yayıncılık.
- Göçer, A. (2021). Türkiye'de dil öğretimi ve ders kitapları. *Türkiye Araştırmaları Literatür Dergisi*, 19(38), 211-232.
- Gümüş, S. (2023). Yabancı dil öğretim materyallerinde kültürel unsurlar. *Dil Dergisi*, 174(1), 7-25.
- Haryanti, M. (2013). *Textbook evaluation in a Private Elementary School in Cimahi*. A Research Paper. Bandung: Indonesia University of Education.
- Işık, B., & Işık, A. (2017). Yabancı Dile Neden Hâlâ Yabancıyız? *Eğitime Bakış* (39), s. 3-11.
- İlhan, M. (2016). A Comparison of the Ability Estimations of Classical Test Theory and the Many Facet Rasch Model in Measurements with Open-ended Questions. *Hacettepe University Journal of Education*. DOI: 10.16986/HUJE.2016015182.
- İşisağ, K. U. (2020). Yabancı dil öğretiminde materyal tasarımı ve çeşitlendirme. *Ankara Üniversitesi Eğitim Bilimleri Fakültesi Dergisi*, 53(1), 273-298.
- Kachru, B. B. (2005). *Asian Englishes: Beyond the Canon*. Hong Kong: Hong Kong University Press.
- Karaman, P. (2021). Türkiye'de dil öğretiminde ölçme ve değerlendirme uygulamaları. *Ölçme ve Değerlendirme Dergisi*, 12(1), 122-141.

- Kırkgöz, Y. (2009). Evaluating the English textbooks for young learners of English at Turkish primary education. *Procedia Social and Behavioral Sciences* 1, 79–83
- Kiely R. (2009). Small answers to the big question: Learning from language program evaluation. *Language Teaching Research*, 13(1), 99-116. doi:10.1177/1362168808095525.
- Linacre, J. M. (2007). *A User's Guide to FACETS: Rasch Model Computer Programs*. Chicago, IL.
- Lunz, M. E., Wright, B. D. & Linacre, J. M. (1990). Measuring the Impact of Judge Severity on Examination Scores. *Applied Measurement in Education*. 3, 4, 331-345.
- Mirici, İ.Hakkı. (2003). The ELT Curricula of the Elementary Education Grades 4 and 5 in Turkey. *Egypt TESOL Journal*. Vol.2, Number 1, Spring 2003, Cairo: Egypt Tesol, ISBN 977-5125-00-X
- Nakamura, N. (2002). Teacher and Peer Assessment. *Educational Studies*. 44, 143, 204-215.
- Polat, M., (2018). Akademisyenlerin Yabancı Dil Bilgisi Seviye Tespit Sınavı'nın (YDS) Geçerliliğiyle İlgili Tutumlarının Belirlenmesi. *Anadolu Üniversitesi Sosyal Bilimler Dergisi*, Vol.18, No.4, 1-28.
- Polat, M., (2020). YDS ve YÖKDİL Sınavlarının Yabancı Dil Eğitimi Üzerine Etkileri. *Gazi Üniversitesi Kastamonu Eğitim Dergisi*, Vol.28, No.5, 2188-2202.
- Polat, M., Turhan, N. S., & TORAMAN, Ç., (2022). Comparison of Classical Test Theory vs. Multi-Facet Rasch Theory. *Pegem Eğitim ve Öğretim Dergisi*, Vol.12, No.2, 213-225.
- Randall, J. & Engelhard, G. Jr. (2009). Examining Teacher Grades Using Rasch Measurement Theory. *Journal of Educational Measurement*, 46, 1, 1-18.
- Sarıçoban, A. & Öztürk, G. (2024). Türkiye'de İngilizce Öğretiminde Materyal Geliştirme ve Değerlendirme. *Hacettepe Üniversitesi Eğitim Fakültesi Dergisi*, 39(1), 214-232.
- Semerçi, Ç. (2011). Mikro Öğretim Uygulamalarının Çok-Yüzeysel Rasch Ölçme Modeli ile Analizi. *Eğitim ve Bilim/ Education and Science*, 36 (161), 14-25.
- Solak, E. (2011). İlköğretim Altı, Yedi Ve Sekizinci Sınıf İngilizce Ders Kitabı Spot On Ders Kitabı Serisinin Değerlendirme Ölçütlerine Göre Değerlendirilmesi. *e-Journal of New World Sciences Academy*, (6),1.
- Stufflebeam, D.L. (1972). *Educational evaluation and decision making*. Illinois: F.E. Peacock Publishers Inc. Itasca.
- Şahin, M. & Yıldırım, D. (2022). Uzman Değerlendirmelerinin Ders Kitapları Revizyonuna Etkisi: MEB İngilizce Kitapları Örneği. *Uluslararası Eğitim Araştırmaları Dergisi*, 13(4), 89-112.
- Tashakkori, A. & Creswell, J. W., (2008). Mixed methodology across disciplines. *Journal of Mixed Methods Research*, 2(1), 3-6.
- T.C. Millî Eğitim Bakanlığı (2023)a. *Türkiye Yüzyılı Eğitim Vizyonu*. Ankara: MEB Yayınları.
- T.C. Millî Eğitim Bakanlığı (2023)b. *İngilizce Dersi Öğretim Programı (İlkokul ve Ortaokul 2-8. Sınıflar)*. Ankara: MEB Yayınları.
- Tomlinson, B. (2003) *Developing Materials for Language Teaching*. London: Continuum Press.
- Türkiye Maarif Foundation (2023). *Değerler Eğitimi Modeli*. İstanbul: TMV Yayınları.
- Yıldırım, A., & Şimşek, H. (2023). *Nitel Araştırma Yöntemleri* (14. Baskı). Ankara: Seçkin Yayıncılık.
- YÖK (2023). *Yabancı Dil Eğitimi Çalıştay Raporu*. Ankara: Yükseköğretim Kurulu Yayınları.

APPENDIX

Dear Participant, this rubric has been prepared for the evaluation of the 10th Grade English Textbook, which you have previously reviewed and agreed to provide your opinions on. Please rate the following statements based on your own opinions considering the book's content.

Thanks for your participation and time.

10th Grade English Textbook Evaluation Rubric					
		Not at all	Partially	Yes	Note
1	The textbook content aligns with the learning outcomes.				
2	The vocabulary is appropriate for the student's level.				
3	The content is organized from simple to complex.				
4	It contains a content that fosters a sense of national unity.				
5	The spelling complies with grammatical rules.				
6	The activities encourage communication in English.				
7	It is sufficient to realise the English course objectives.				
8	It will encourage students to learn outside the classroom.				
9	It contains activities from various other cultures.				
10	It contains content that reflects Turkish family.				
11	The printing quality of the textbook is sufficient.				
12	It contains texts that explain Türkiye's global role.				
13	The variety of topics in the textbook is sufficient.				
14	The audio quality of the recordings is understandable.				
15	It will increase students' interest in learning English.				