

# Development of a Self-Efficacy Scale for Arabic Reading Comprehension Among High School Students

Sara Akkuş<sup>1</sup> 

## Abstract

This study aimed to develop a self-efficacy scale to assess high school students' perceived competence in Arabic reading comprehension. Accordingly, a draft version of the "High School Students' Self-Efficacy Scale for Arabic Reading Comprehension" was constructed as a 5-point Likert-type instrument. Following expert evaluations ( $n=8$ ), the initial pool of 46 items was reduced to a 23-item trial form based on content validity analysis. Exploratory factor analysis (EFA) was conducted with data from 210 students, while confirmatory factor analysis (CFA) was performed with a separate sample of 473 students. In this study, EFA and CFA were performed on separate samples. While this is methodologically appropriate, it may have constrained the model's fit. Moreover, the use of an all-male sample for the EFA presents a potential limitation in terms of demographic diversity. EFA results indicated a unidimensional structure. The CFA results demonstrated acceptable model fit indices ( $RMSEA = 0.06$ ,  $\chi^2/df = 2.59$ ,  $GFI = 0.87$ ,  $CFI = 0.95$ ), confirming the construct validity of the scale. The internal consistency of the scale was found to be high, with a Cronbach's alpha coefficient of .957. Although the scale demonstrated strong psychometric properties, the absence of gender diversity in the EFA sample and the lack of subsequent invariance testing are acknowledged as limitations. Nevertheless, despite these constraints, the developed scale represents a valuable tool for educators, researchers, and curriculum developers aiming to assess high school students' self-efficacy in Arabic reading comprehension.

## Keywords

self-efficacy, scale development, validity and reliability, Arabic reading comprehension, Arabic language teaching

## Introduction

Within the framework of social-cognitive theory, the concept of self-efficacy was first introduced by Bandura and thus incorporated into the literature (Zimmerman, 2002). Self-efficacy is defined as "an individual's self-perception of their capacity to carry out tasks required in a given situation" (Bandura, 1997; McFall, 1982). Zimmerman (2002), on the other hand, conceptualizes it as "beliefs that shape a person's emotions, thoughts, motivations, and patterns of behavior." An individual's perception of self-efficacy influences not only their behaviors but also their levels of achievement and motivation (Bandura, 1997; Zimmerman, 2002). It has been stated that individuals with high self-efficacy beliefs tend to set more ambitious goals and exert greater effort to achieve them (Warner & Schwarzer, 2017). Such individuals generally attribute failures to external factors or insufficient effort, thereby maintaining their belief that they will

succeed in subsequent attempts. Conversely, individuals with low self-efficacy beliefs are more inclined to exert less effort and abandon their goals, which often leads to repeated failure.

According to Leithwood (2007), self-efficacy beliefs are not a direct reflection of an individual's actual performance or abilities. Rather, self-efficacy is a perception grounded in personal judgments and self-assessments. For instance, even if an individual holds strong self-efficacy beliefs regarding a particular skill, their actual

<sup>1</sup>Fatih Sultan Mehmet Vakif University, İstanbul, Turkey

### Corresponding Author:

Sara Akkuş, Fatih Sultan Mehmet Vakif University, Valide-i Atik Mah. Eski Toptası Cad. No. 93 Atik Valide Külliyesi, Üsküdar, İstanbul 34083, Turkey.  
Email: sakkus@fsm.edu.tr

Data Availability Statement included at the end of the article



performance may fall short of what they perceive (Leithwood, 2007). In this context, the concept of self-efficacy is regarded as a form of “perceived reality” (Y. Wang & Sun, 2024). Nonetheless, Bandura (1997) emphasizes that the perception of self-efficacy is a determining factor in how individuals employ their skills and capacities. According to Bandura, as self-efficacy beliefs strengthen, the effort and perseverance exhibited by individuals increase accordingly. Thus, self-efficacy beliefs are considered a significant mechanism guiding the process of behavioral change.

Bandura (1997) argues that an individual’s self-efficacy belief is influenced by four primary sources. These sources play a critical role in either enhancing or diminishing self-confidence and shape self-efficacy perceptions depending on how individuals interpret their environmental and internal interactions. Mastery experiences are among the strongest determinants of self-efficacy beliefs. Positive outcomes achieved in the past reinforce individuals’ confidence that they will succeed in similar future situations (Y. Wang & Sun, 2024). Vicarious experiences, derived from observing the success of others, may foster the perception that one can achieve similar success. This effect is particularly strengthened when the observer shares similar characteristics with the model (Bandura, 1997). Social persuasion is another important tool for enhancing self-belief, although its influence may be more limited compared to other sources (Y. Wang & Sun, 2024). Physiological and psychological states also shape perceptions of self-efficacy. For instance, stress, anxiety, or fatigue may weaken self-efficacy beliefs, while a positive mood and physical well-being may strengthen them (Zimmerman, 2002).

In Bandura’s (1997) social-cognitive theory, self-efficacy is defined as a critical factor that influences both motivation and academic performance. Individuals with high self-efficacy set more ambitious goals during the learning process, exert greater effort to achieve them, and demonstrate resilience when facing obstacles (Zimmerman, 2002). Self-efficacy beliefs influence behaviors directly related to academic success, such as the choice of learning strategies, attention control, and problem-solving. Students with strong self-efficacy beliefs employ more effective strategies during the learning process, which facilitates access to knowledge and transforms it into long-term learning. In contrast, when self-efficacy is low, students typically adopt more passive learning strategies and are more prone to abandoning the learning process.

Foreign language learning is generally a long-term process requiring persistence, determination, and the use of effective learning strategies. Within this process, individuals’ perceptions of self-efficacy emerge as one of the

most significant psychological factors determining success in language acquisition. In foreign language learning, self-efficacy directly affects motivation levels. Learners with high self-efficacy are more determined and motivated to cope with challenges encountered during the language learning process. For example, when faced with tasks such as mastering a difficult grammar rule or participating in a speaking activity, individuals with high self-efficacy are more likely to exert effort and feel more willing to engage in these tasks (Dörnyei, 2005). Conversely, individuals with low self-efficacy are more likely to give up easily and perceive themselves as inadequate in language learning. In their evaluation of the role of self-efficacy in second/foreign language acquisition, Raoofi et al. (2012) reviewed 32 studies. Their findings revealed that self-efficacy perception is a strong predictor of foreign language learning performance. Moreover, it was found to exert a positive effect on core language skills (reading, listening, speaking, and writing) and to be directly associated with lower anxiety and higher motivation.

Therefore, the study aims to develop a reliable and valid self-efficacy scale to measure high school students’ perceived self-efficacy in Arabic reading comprehension.

## Literature Review

The literature has strongly demonstrated that self-efficacy beliefs in foreign language learning have significant effects on student achievement and motivation (Bandura, 1997; Zimmerman, 2002). Studies conducted in Turkey have generally employed quantitative and cross-sectional methods, identifying positive correlations between self-efficacy and academic achievement. In contrast, international studies have taken a more comprehensive approach, using longitudinal and experimental designs to reveal causal relationships (Hsieh & Schallert, 2008; Mills et al., 2006). For instance, Mills et al. (2007) demonstrated that among university-level French learners, self-efficacy beliefs influenced not only academic success but also learning motivation. Additionally, while Graham (2007), Egel (2009), and Magogwe and Oliver (2007) emphasized the effect of learning strategies and learner autonomy on self-efficacy, studies such as C. Wang and Pape (2007) demonstrated that cultural context plays a decisive role in self-efficacy perceptions. More recently, Kim et al. (2024) highlighted how digital learning environments and socio-cultural factors can influence self-efficacy beliefs in language learners. These differences indicate that the international literature addresses cross-cultural variables, individual differences (e.g., age, gender, and social factors), and contextual elements across a much broader spectrum. However, studies conducted in Turkey tend to focus on more

homogeneous samples and narrower contexts. Despite the growing body of research acknowledging sociocultural influences on language learning, many scales development studies-including those conducted in the Turkish context-have insufficiently addressed the sociocultural diversity of learners. Factors such as socioeconomic background, linguistic environment, and cultural identity can significantly shape learners' self-efficacy beliefs (Kim et al., 2024). Limited consideration of these variables may, however, restrict the generalizability of findings. Therefore, this study seeks to contribute to the literature by developing a scale that recognizes this diversity and can be applied across varied learner profiles.

Particularly in less-studied languages such as Arabic, there is a notable lack of valid and reliable measurement tools that assess self-efficacy perceptions in relation to specific language skills, such as reading comprehension. Within this context, the present study emerges as one of the first to apply self-efficacy theory to Arabic reading comprehension, aiming to make a structural contribution to the field through the development of a measurement instrument.

The literature review revealed that only a limited number of national and international scale development studies have sought to measure self-efficacy perceptions in Arabic learning. Among these are the *Arabic Speaking Skills Self-Efficacy Perception Scale* developed by A. Yeşilyurt (2017), the *Arabic Language Efficacy Questionnaire (ALEQ)* designed by Mustapha et al. (2013), and the *Arabic Learning Motivation Scale (ALMS)* introduced by Düzgün and Kırkıç (2023). This suggests the need for further scale development studies specifically focused on this field. The existing literature reveals that scale development studies aimed at assessing self-efficacy perceptions in Arabic language learning are quite limited. The *Arabic Speaking Skills Self-Efficacy Perception Scale* developed by A. Yeşilyurt (2017) focuses exclusively on oral skills, thereby failing to fully capture the multidimensional nature of the self-efficacy construct. The ALEQ, presented by Mustapha et al. (2013), addresses Arabic language proficiency more generally, yet it does not position the concept of self-efficacy at its core. Meanwhile, the ALMS developed by Düzgün and Kırkıç (2023) explores a different construct by emphasizing motivation. Although these three studies offer valuable measurement tools related to Arabic language learning, they collectively highlight the ongoing absence of a comprehensive and holistic scale that assesses psychological constructs-such as self-efficacy-that directly influence the language learning process. This gap is particularly significant in contexts where Arabic is taught as a foreign language, where there remains a pressing need for valid and reliable instruments to evaluate learners' beliefs about their own

learning capabilities. In this regard, the present study not only seeks to address this gap but also aspires to contribute to a deeper understanding of the language learning processes experienced by individuals learning Arabic.

Research on self-efficacy beliefs in foreign language classes has generally shown a positive relationship between self-efficacy beliefs and student achievement, highlighting the critical role of self-efficacy in attaining success. However, the lack of an instrument specifically designed to examine the self-efficacy beliefs of İmam-Hatip high school students regarding Arabic reading comprehension is noteworthy. At a time when Arabic learning is gaining increasing importance, there is a clear need to develop scientifically grounded tools to assess self-efficacy beliefs in this area.

In light of this importance, the present study aims to develop a reliable and valid scale to determine İmam-Hatip high school students' self-efficacy perceptions of Arabic reading comprehension skills. It is anticipated that such a scale will not only help identify students' self-efficacy levels but also assist Arabic teachers in planning their lessons more effectively and enable school administrations to create learning environments that support students' self-efficacy beliefs. Furthermore, the absence of a reliable and valid measurement tool on Arabic reading comprehension self-efficacy in Turkey, together with the limited research on this subject, underscores the necessity of this scale to fill the existing gap in the literature.

## Method

### Design and Participants

This study aimed to develop a measurement tool that would allow for the valid and reliable interpretation of self-efficacy scores related to Arabic reading comprehension skills among İmam-Hatip high school students. Accordingly, the study was designed in line with the descriptive survey model, one of the quantitative research methods. The descriptive survey model encompasses the process of collecting and analyzing data to describe the existing characteristics of individuals (Creswell & Creswell, 2018; Morgado et al., 2018). Within this framework, a pilot study of the scale was conducted with an appropriate sample group to examine the psychometric properties of the instrument. The data obtained from the pilot application were used for validity and reliability analyses of the scale.

The sample for this study was drawn from three different schools using the easily accessible sampling technique, a type of purposive sampling method. The scale development process was carried out in two stages: in the first stage, exploratory factor analysis (EFA) was applied, and in the second stage, confirmatory factor

**Table 1.** Information about the Research Sample (N = 683).

Application type	Grade level	N	Percentage (%)
Exploratory factor analysis (School 1)	Grade 9	77	11.3
	Grade 10	49	7.2
	Grade 11	61	8.9
	Grade 12	23	3.4
Confirmatory factor analysis (School 2)	Grade 9	105	15.4
	Grade 10	116	17.0
Confirmatory factor analysis (School 3)	Grade 9	101	14.8
	Grade 10	105	15.4
	Grade 11	46	6.7
Total		683	100

analysis (CFA) was performed. Information about the sample is presented in Table 1.

As shown in Table 1, 210 male students from School 1 participated in the exploratory factor analysis (EFA). In the subsequent stage, to confirm the results, 221 female students from School 2 and 252 male students from School 3 completed the scale for the confirmatory factor analysis (CFA;  $n = 473$ ). Students from grades 9, 10, 11, and 12 were included in the sample, indicating that all levels of secondary education were represented.

The scale development process was conducted in two stages: exploratory factor analysis (EFA) and confirmatory factor analysis (CFA). In the EFA stage, a sample consisting solely of male students was used, whereas in the CFA stage, a more heterogeneous group in terms of gender (both male and female students) participated. This distribution was shaped by the existing student profiles of the imam-hatip schools (being either all-male or all-female institutions) and accessibility conditions. In the scale development literature, it is common and acceptable to conduct EFA and CFA with different samples. This approach is important for testing the consistency of the scale across different groups. In the present study, a gender-diverse sample was employed in the CFA stage to examine the scale's performance across various demographic structures. However, the differences between the EFA and CFA samples may have limited model fit, and this has been considered one of the study's limitations.

## Data Collection

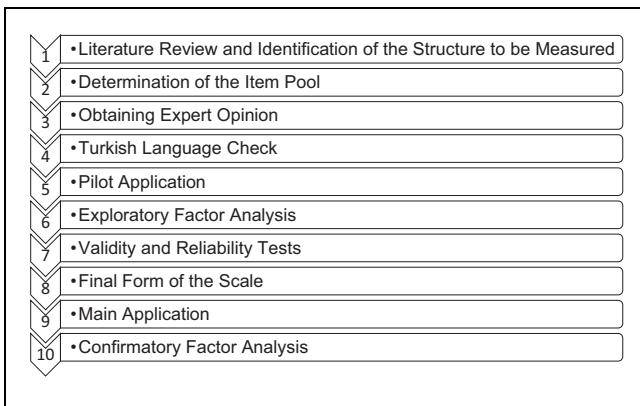
The scale development process involves systematically executed steps designed to measure a specific construct. The construct targeted in this study is secondary school students' self-efficacy perceptions regarding Arabic reading comprehension. The concept of self-efficacy, structured on the basis of Bandura's (1997) social-cognitive theory, refers to an individual's belief in their capacity to accomplish a specific task successfully. Self-efficacy is a crucial psychological variable that directly influences an

individual's motivation, effort in learning, and learning strategies. Therefore, in foreign language learning—particularly in the acquisition of a language with distinct structural features such as Arabic—measuring students' self-efficacy perceptions is considered significant both for the instructional process and for predicting achievement. In this context, the scope of the construct to be measured was defined by drawing upon both theoretical and practical foundations. To this end, the reading comprehension outcomes included in the European Language Portfolio (Council of Europe, 2001) and the Arabic Teaching Curriculum of the Ministry of National Education (MoNE, 2022) were thoroughly examined, and it was determined that a measurement tool was needed to assess the extent to which students achieve these skills. In addition, national and international literature on the relationship between self-efficacy and language skills was reviewed, emphasizing that self-efficacy has a decisive influence on individuals' behavioral performance (Raoofi et al., 2012; Zimmerman, 2002). Subsequently, the items were reviewed by a Turkish language specialist to ensure linguistic accuracy and clarity.

The scale development process was carried out in two stages following the creation of the item pool and preliminary testing. In the first stage, the factor structure was explored through exploratory factor analysis (EFA) conducted on a pilot sample; in the second stage, confirmatory factor analysis (CFA) was applied to test the validity of the obtained model. In the literature, EFA is recognized as an initial step in testing construct validity; however, verifying this structure with CFA in different samples is acknowledged to both strengthen validity evidence and enhance generalizability (Aslan et al., 2025). Creswell and Creswell (2018) further emphasizes that supporting a factor structure identified by EFA with CFA in another sample consolidates the structural consistency of the scale and provides an opportunity to test its reliability across groups with different demographic characteristics. The planning of the study carried out to develop a self-efficacy scale for Arabic reading comprehension is presented below, with details provided in the subsequent sections (Figure 1).

## Defining the Construct and Literature Review

Defining the construct constitutes a fundamental and critical stage in the scale development process. At this stage, it is essential to accurately specify the characteristics, dimensions, and boundaries of the concept to be measured. Determining the scope of the construct comprehensively and on a scientific basis is vital for the validity and reliability of the results derived from the scale. In this regard, Kline (2016) considers the clear definition of construct characteristics as the most crucial



**Figure 1.** Scale development process.

Source. DeVellis (2012).

step in the scale development process. Through an extensive literature review and expert consultation, the theoretical framework of the concept to be measured was grounded on a robust basis, thereby ensuring the development of a valid measurement instrument. The construct in this study is secondary school students' self-efficacy perceptions regarding Arabic reading comprehension. Based on Bandura's (1997) social-cognitive theory, self-efficacy refers to an individual's belief in their ability to carry out a specific task successfully. As a key psychological variable, self-efficacy directly influences motivation, learning effort, and strategy use. In this sense, measuring students' self-efficacy perceptions in foreign language learning—particularly in the case of Arabic with its unique linguistic structures—is considered highly significant for both instructional practice and predicting achievement. Accordingly, the scope of the construct was determined using both theoretical and practical foundations. Specifically, the reading comprehension outcomes in the European Language Portfolio and the MoNE Arabic Teaching Curriculum were thoroughly examined, which revealed the need for a measurement tool assessing the degree to which students achieve these competencies. Additionally, local and international studies demonstrating the relationship between self-efficacy and language skills were reviewed, highlighting that self-efficacy exerts a decisive effect on behavioral performance (Raoofi et al., 2012; Zimmerman, 2002). In line with the literature and the theoretical framework, the measurement targeted students' self-efficacy perceptions regarding sub-skills such as comprehending, analyzing, making inferences from Arabic texts, predicting word meanings from context, and interpreting graphics and visuals. The scope of the construct was structured to reflect not only the cognitive dimension but also students' confidence in their abilities. Thus, the scale aimed to reveal not only students' knowledge levels but also

their belief in their capacity to use that knowledge. In delineating the theoretical boundaries of the construct, previously developed scales in the field of Arabic teaching (S. Yeşilyurt & Çapraz, 2018) were also examined. However, these instruments were found to focus on speaking or general language proficiency and were limited in measuring self-efficacy in reading comprehension. Therefore, the construct was elaborated to address a specific measurement need in a previously underexplored area, thus constituting the study's original contribution. Consequently, Arabic reading comprehension self-efficacy was conceptualized holistically at the conceptual, contextual, and practical levels, encompassing students' abilities to understand, interpret, extract, and apply information from texts, as well as their confidence in these processes.

### **Item Pool Development**

To better define the scope and characteristics of the construct, both national and international standards were utilized. Particular attention was paid to the reading comprehension outcomes included in the European Language Portfolio and the MoNE Arabic Teaching Curriculum. Furthermore, items from similar scales identified through literature review were examined in detail (Dang, 2024; Güngör & Kan, 2020; Kim et al., 2024; Kurudayioğlu, Yazıcı & Göktentürk, 2021; Sağlam & Arslan, 2018; A. Yeşilyurt, 2017). This review enriched the theoretical framework and provided a model structure to guide item development. In line with recommendations in the literature, special care was taken to generate a large number of items when shaping the item pool (DeVellis, 2012; Kline, 2016). Additionally, internal consistency within each item group was prioritized (Kline, 2016). In item writing, the psychosocial characteristics of the target group were considered, and the language of the items was simplified accordingly.

### **Expert Review and Language Check**

DeVellis (2012) defines content validity as a type of validity established by evaluating a measurement instrument based on expert opinions. Content validity is typically assessed with the contributions of experts in the relevant field. To ensure systematic expert feedback, the researcher developed a form including all 46 items with the options "Appropriate," "Appropriate/Insufficient," and "Not Appropriate." Opinions were collected from two measurement and evaluation experts, one Turkish linguist, and five language education specialists. To obtain reliable results from these evaluations, the Lawshe method was applied (S. Yeşilyurt & Çapraz, 2018). Accordingly, the content validity ratio (CVR) was

**Table 2.** Skewness-Kurtosis Test Results.

Variable	N	Skewness		Kurtosis	
	Statistic	Statistic	Standard error	Statistic	Standard error
Scale total score	210	-0.554	0.168	0.414	0.334

calculated using the formula  $CVR = (Ne - N/2)/(N/2)$ , where  $Ne$  represents the number of experts rating the item as “appropriate.” Based on the critical value established by Shrotryia and Dhanda (2019), when the total number of experts is eight, the critical value is 0.750. Consequently, 21 items falling below this threshold were excluded from the scale. For the remaining 25 items, the content validity index (CVI) calculation yielded an average of 0.778, which indicates that the scale provided sufficient evidence of content validity. Subsequently, instructions and demographic information were added to the scale, rendering it ready for the first administration.

### Ethical Considerations

To conduct the study, approval was first obtained from the Social and Human Sciences Research Ethics Committee of İstanbul University (Date and Reference: 14.02.2024/2419875). Subsequently, with the permission granted by the İstanbul Provincial Directorate of National Education (Document No: 99467957, dated 25.03.2024), the scale was administered to students enrolled in imam-hatip high schools.

### Results and Discussion

The *Self-Efficacy Scale for Secondary School Students’ Arabic Reading Comprehension Skills* was administered as a pilot study to 210 students enrolled in Arabic courses in grades 9, 10, 11, and 12 at *School 1*. The collected data were first entered and organized in Excel, then transferred to IBM SPSS Statistics 21. Since the scale did not include any negatively worded items requiring reverse coding, no reverse coding process was performed. A frequency analysis was then conducted, confirming that there were no missing data. Skewness and kurtosis values were examined to determine whether the data were normally distributed (Table 2).

According to Tabachnick and Fidell (2019), skewness and kurtosis values should fall between -1.5 and +1.5. Based on these statistics, the total score distribution of the scale was found to be very close to normal. Since skewness and kurtosis values were within acceptable limits, the assumption of normality was confirmed. These results indicated that parametric tests were applicable.

**Table 3.** KMO and Bartlett’s Test Results for the Self-Efficacy Scale for Secondary School Students’ Arabic Reading Comprehension Skills.

Test	Statistic	Value
Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy	KMO	0.951
Bartlett Sphericity test	Chi-square ( $\chi^2$ ) Degrees of freedom (df) Significance ( $p$ )	2,972.941 253 .00

$p < .05$ .

### Exploratory Factor Analysis

Exploratory Factor Analysis is conducted to identify the underlying factors or dimensions of a set of observed variables by analyzing their interrelationships (Morgado et al., 2018). This method plays a critical role in scale development, validity analysis, and the generation of hypotheses concerning theoretical structures. In the results of the analysis, the Kaiser-Meyer-Olkin (KMO) test was examined first. The KMO value should not fall below 0.50, and a range of 0.60 to 0.70 is expected (Morgado et al., 2018).

As shown in Table 3, the Kaiser-Meyer-Olkin (KMO) value was found to be suitable (KMO = 0.951). This indicated that the sample size was appropriate for factor analysis. Bartlett’s Test of Sphericity was also found to be significant ( $\chi^2 = 2,972.941$ ,  $df = 253$ ,  $p < .00$ ), confirming the suitability of the data structure for factor analysis. Since the KMO value (KMO = 0.951) did not fall below 0.50 (Table 4), anti-image matrix results were not considered (Table 4).

As shown in Table 4, the scale included a single factor with an eigenvalue greater than 1, along with its explained variance percentage. Accordingly, the first factor, with an eigenvalue of 11.633, accounted for 50.578% of the total variance. A total variance explanation rate between 40% and 60% is considered acceptable (Pallant, 2020).

In scale development, factor loadings represent the coefficient of the relationship between an item and a given sub-dimension. The literature generally accepts factor loadings in the range of 0.30 to 0.40 as sufficient for identifying factor patterns (Morgado et al., 2018). In

**Table 4.** Factor Analysis Results of the Self-Efficacy Scale for Secondary School Students' Arabic Reading Comprehension Skills.

Dimensions	Eigenvalue	Percentage of variance explained (total)	Percentage of variance explained (cumulative)
Total dimension	11.633	50.578	50.578

**Table 5.** Factor Loadings of the Items in the Self-Efficacy Scale for Secondary School Students' Arabic Reading Comprehension Skills.

Item	Factor loading
i12 I can understand basic patterns related to myself and my environment by reading them in a fluent conversational text.	0.786
i3 I can visualize what I read in an Arabic text in my mind.	0.778
i17 I can distinguish place/time/person elements in the Arabic text I read.	0.774
i2 I can easily understand when I read an Arabic text without pictures.	0.773
i15 I can establish cause-and-effect relationships in the Arabic text I read.	0.764
i5 I can distinguish comparative expressions in the Arabic text I read.	0.763
i1 I can generally understand the main idea when I read an Arabic text.	0.758
i19 I can easily understand when I read an Arabic short story.	0.749
i13 I can read and comprehend the basic idea in Arabic short messages and announcements.	0.748
i11 I can understand emotional emphases in the Arabic text I read.	0.748
i7 I can answer questions about the Arabic text I read.	0.738
i16 I can ask questions about the Arabic text I read.	0.737
i9 I can match concepts and visuals based on the Arabic text I read.	0.732
i14 I can find the information I am looking for in an Arabic text.	0.726
i20 I can understand when I read Arabic directions.	0.713
i22 I can conduct Arabic research on a given simple topic.	0.701
i23 I can scan Arabic sources on a given simple topic.	0.688
i10 I can guess the meanings of new words based on clues given in the Arabic text.	0.677
i8 I can distinguish conjunctions/prepositions from other sentence elements in the Arabic text I read.	0.642
i25 I can interpret simply prepared Arabic graphics, tables, and other visuals at a basic level.	0.637
i21 I can use correct diacritical marks while reading an Arabic text.	0.587
i4 I can distinguish numerical expressions from other expressions in the Arabic text I read.	0.551
i24 I can understand directives such as "come, go, sit, stand" in the Arabic text I read.	0.495
i6 I can understand time expressions (past tense, present tense, future tense, etc.) in the Arabic text I read.	0.291
i18 I can understand when I read an Arabic dialogue text.	0.289

in this study, a cutoff value of 0.30 was adopted. The factor loadings for the single-dimension structure of the scale, following all analyses and item removals, are presented in Table 5.

Table 5 presents the factor loadings of the items in the overall dimension. The majority of items demonstrated strong factor loadings (0.70 and above). This indicates that the scale exhibited a consistent structure regarding the self-efficacy construct it aimed to measure. The highest factor loading was observed in item i12 (0.786). This item—"I can read and understand basic expressions related to myself and my environment in a fluent Arabic text"—provides an important indicator of participants' self-efficacy in understanding fundamental information in Arabic texts. Other high factor loadings were concentrated in items such as i3 (0.778), i17 (0.774), i2 (0.773), i15 (0.764), and i5 (0.763).

Based on the factor loadings obtained from exploratory factor analysis (EFA), items i6 ("I can understand

the time expressions (past, present, future, etc.) in an Arabic text I read") and i18 ("I can understand when I read an Arabic dialogue text") had factor loadings of 0.291 and 0.289, respectively, falling below the 0.30 threshold. According to widely accepted approaches in the literature, items with factor loadings below 0.30 should be excluded from the scale, as they fail to adequately represent the intended construct (Morgado et al., 2018). Therefore, items i6 and i18 were removed. This removal was carried out to enhance the construct validity and reliability of the scale. These results demonstrate that the scale has meaningful construct validity and that the dimensions sufficiently represent student self-efficacy.

## Reliability Tests of the Pilot Study

For a test to accurately measure the intended construct, it must first demonstrate reliability (DeVellis, 2012). The internal consistency method, particularly

**Table 6.** Internal Consistency Coefficients of the Self-Efficacy Scale for Secondary School Students' Arabic Reading Comprehension Skills.

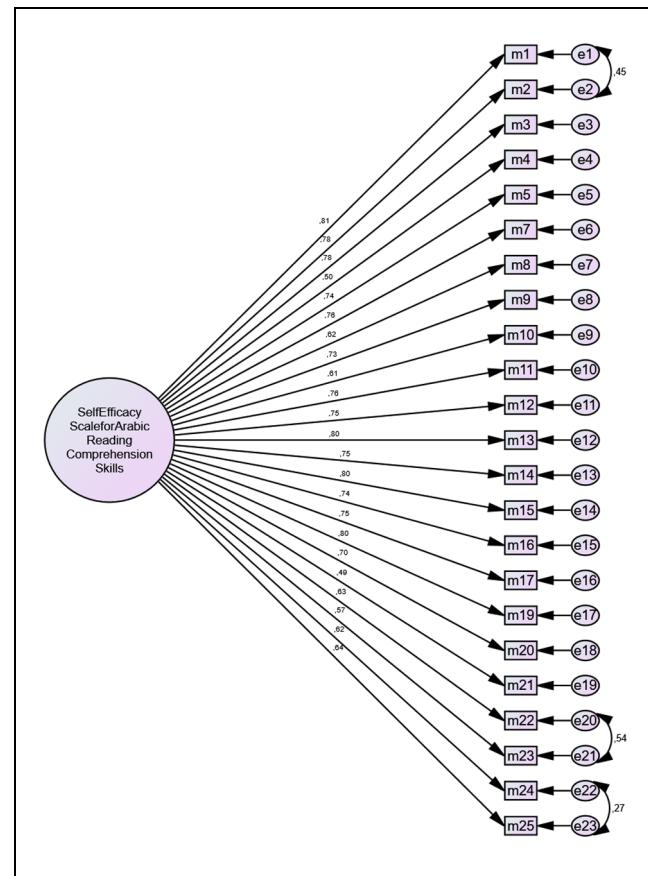
Coefficients	<i>n</i>	<i>r</i>
Cronbach $\alpha$	210	.96
Spearman-Brown	210	.93
Guttman's Lambda-2	210	1.92

Cronbach's alpha coefficient, is commonly used to assess reliability in Likert-type scales. Using the collected data, the internal consistency coefficient of the scale was calculated, and the reliability results are presented below.

As shown in Table 6, the reliability coefficient was calculated as  $\alpha = .96$ . According to Boateng et al. (2018), the reliability coefficient should exceed  $\alpha > .70$ , while values in the range of  $.80 \leq \alpha < 1.00$  are considered highly reliable. The Spearman-Brown coefficient, calculated using the split-half method, indicated that the two halves of the scale contributed consistently ( $S = .93$ ). The Lambda-2 coefficient, which serves as an alternative indicator of item consistency, also yielded a very high value, confirming the excellent reliability of the scale ( $G = 1.92$ ). In conclusion, all items in the draft scale were found to measure the same construct and demonstrated a high level of reliability.

## Confirmatory Factor Analysis

To further test the reliability and validity of the scale, which was shown to have a single sub-dimensional structure, CFA was conducted on a different sample. Confirmatory factor analysis is an advanced statistical method used to test whether a scale or measurement model conforms to a previously established theoretical structure (Tabachnick & Fidell, 2019). In this context, the 23-item scale was administered to 473 secondary school students, and CFA was conducted using the collected data. Fit indices ( $\chi^2/SD$ , AGFI, GFI, NFI, CFI, IFI, TLI, RMSEA) were calculated to confirm the three-factor structure of the scale. Acceptable threshold values were set as follows: CFI/TLI  $\geq 0.90$ ; RMSEA  $\leq 0.08$ ; SRMR  $\leq 0.08$  (Hu & Bentler, 1999; Kline, 2016). Inter-factor correlations were analyzed to examine the independence or interrelatedness of the sub-dimensions. Standardized beta coefficients were first examined and are presented in Figure 2. When the path diagram of the confirmatory factor analysis shown in Figure 2 is examined, the factor loadings of the items represent the relationships between each observable variable ( $i_1, i_2, \dots, i_{25}$ ) and the latent variable (overall dimension). Factor loadings generally range between 0 and 1, with higher



**Figure 2.** Path diagram of the confirmatory factor analysis (CFA) results of the self-efficacy scale for secondary school students' Arabic reading comprehension skills.

values indicating stronger relationships (Tabachnick & Fidell, 2019). Specifically, loadings of 0.70 and above suggest a strong relationship, 0.50 to 0.70 a moderate relationship, 0.30 to 0.50 a weak relationship, and below 0.30 a very weak relationship. In the diagram, the distribution of item factor loadings appears to be generally satisfactory.

Covariances were identified between the error terms  $e_1$  and  $e_2$ ,  $e_{20}$  and  $e_{21}$ , and  $e_{22}$  and  $e_{23}$ . These correlations were retained because they reflect shared measurement variance resulting from conceptual and linguistic similarities between the related items rather than methodological bias. For example, the covariance between  $e_{20}$  and  $e_{21}$  likely stems from the conceptual proximity of the items "I can conduct Arabic research on a given simple topic" and "I can scan Arabic sources on a given simple topic" both of which involve similar cognitive processes related to information retrieval and comprehension in Arabic texts. Likewise, the correlation between  $e_{22}$  and  $e_{23}$  arises from their shared focus on interpreting visual or directive elements within reading tasks. In the context of confirmatory factor analysis,

**Table 7.** Unstandardized Regression Coefficients of the Items.

Item			Std estimate	SE	CR	p
i1	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.81			
i2	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.78	0.04	26.81	***
i3	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.78	0.05	19.66	***
i4	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.50	0.06	11.34	***
i5	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.74	0.05	18.50	***
i7	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.76	0.05	18.93	***
i8	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.62	0.06	14.49	***
i9	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.73	0.05	17.93	***
i10	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.61	0.05	14.30	***
i11	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.76	0.05	19.17	***
i12	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.75	0.05	18.65	***
i13	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.80	0.05	20.63	***
i14	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.75	0.05	18.82	***
i15	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.80	0.05	20.39	***
i16	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.74	0.05	18.22	***
i17	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.75	0.05	18.74	***
i19	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.80	0.05	20.58	***
i20	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.70	0.05	17.02	***
i21	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.49	0.06	11.16	***
i22	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.63	0.06	14.94	***
i23	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.57	0.05	13.17	***
i24	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.62	0.06	14.71	***
i25	<—	Self-Efficacy Scale for Arabic Reading Comprehension Skills	0.64	0.06	15.20	***

\*\*\*p &lt; .001.

**Table 8.** Critical Values of Goodness-of-Fit Indices for the Model (Hu & Bentler, 1999; Kline, 2016).

Fit indices	Good fit	Acceptable fit	Research finding	Fit level
$\chi^2/SD$	<3	<5	2.59	Good fit
AGFI	≥ 0.90	≥ 0.80	0.89	Acceptable fit
GFI	≥ 0.90	≥ 0.85	0.87	Acceptable fit
NFI	≥ 0.95	≥ 0.90	0.92	Poor fit
CFI	≥ 0.95	≥ 0.90	0.95	Good fit
IFI	≥ 0.95	≥ 0.90	0.95	Good fit
TLI	≥ 0.95	≥ 0.90	0.94	Acceptable fit
RMSEA	≤ 0.05	≤ 0.08	0.06	Acceptable fit

allowing theoretically justified error correlations is considered acceptable when such relationships reflect conceptually related subcomponents within the same construct (Kline, 2016). Furthermore, these covariances contributed to improving the model's overall fit indices ( $RMSEA = 0.06$ ,  $\chi^2/df = 2.59$ ,  $CFI = 0.95$ ) without altering the theoretical integrity of the model. However, recognizing that excessive error correlations may suggest redundancy among items, future studies are encouraged to reassess these relationships using broader and more diverse samples to confirm whether the observed covariances represent genuine conceptual overlap or sample-specific variation.

The unstandardized regression coefficients presented in Table 7 reveal the extent to which each item represents the latent construct being measured. The fact that all coefficients were significant ( $p < .001$ ) indicates that the overall model contains meaningful relationships (Kline, 2016). This finding points to a model that meets the fundamental assumptions of confirmatory factor analysis (Brown, 2015).

Examination of the regression coefficients showed that items i1 (0.81), i2 (0.78), i3 (0.78), i13 (0.80), i15 (0.80), and i19 (0.80) had high coefficients. On the other hand, items such as i4 (0.50) and i21 (0.49) had relatively lower coefficients but remained acceptable for inclusion in the model, as they exceeded the 0.40 threshold (Tabachnick & Fidell, 2019). These two items were retained to preserve the content validity and conceptual scope of the scale. Specifically, item i14 (I can distinguish numerical expressions from other expressions in the Arabic text I read) and item i21 (I can use correct diacritical marks while reading an Arabic text) represent distinctive aspects of Arabic reading competence that are crucial for understanding text meaning and pronunciation accuracy. Excluding these items would have narrowed the conceptual coverage of the scale and reduced its representation of Arabic-specific reading skills. Furthermore, both items contributed positively to the overall reliability ( $\alpha = .957$ ), supporting their inclusion as methodologically justified. Subsequently, the goodness-of-fit indices of the

**Table 9.** Item Analyses of the CFA Data.

Item	Mean	SD	Item-total correlation	Cronbach's $\alpha$ if item deleted
i1	3.03	1.182	.791	.954
i2	2.51	1.157	.762	.955
i3	2.90	1.221	.754	.955
i5	2.51	1.162	.718	.955
i7	3.01	1.165	.739	.955
i8	2.68	1.317	.602	.956
i9	3.49	1.175	.710	.955
i10	2.94	1.212	.604	.956
i11	2.50	1.175	.741	.955
i12	2.52	1.252	.725	.955
i16	2.66	1.234	.717	.955
i17	3.42	1.253	.736	.955
i19	2.82	1.283	.779	.954
i14	2.89	1.222	.734	.955
i15	2.52	1.187	.770	.954
i20	2.19	1.152	.688	.955
i22	2.38	1.248	.640	.956
i23	2.16	1.150	.582	.956
i13	2.76	1.184	.779	.954
i24	3.84	1.248	.625	.956
i25	3.36	1.260	.648	.956

**Table 10.** Internal Consistency Coefficients of the CFA Data.

Coefficients	CFA sample	r
Cronbach $\alpha$	473	.957
Guttman's Lambda-2	473	.958

measurement model from the CFA were examined and are presented in Table 8.

As shown in Table 8, the initial model fell within acceptable limits regarding goodness-of-fit indices. The  $\chi^2/df$  ratio, which evaluates the fit of the structural equation model by comparing the chi-square value (CMIN) with the degrees of freedom, indicated acceptable fit ( $CMIN/df = 2.59$ ). The model demonstrated a reasonable fit to the data, though not a perfect fit. The adjusted goodness-of-fit index (AGFI), which adjusts the GFI value by accounting for model complexity (Kline, 2016), was 0.89, within the valid range of acceptable fit. The GFI, which reflects how well the model fits the data in structural equation modeling (Hu & Bentler, 1999), was 0.87, observed at the lower boundary of acceptable fit. The normed fit index (NFI), which compares the model fit with that of a null model assuming independence among variables, exceeded the acceptable threshold ( $NFI = 0.92$ ). The comparative fit index (CFI), which evaluates model fit relative to the null model, indicated an acceptable level of fit ( $CFI = 0.95$ ). Another index used in confirmatory factor analysis, the incremental fit

index (IFI), was also at an acceptable level ( $IFI = 0.95$ ), supporting the adequacy of the model. The Tucker-Lewis Index (TLI), which accounts for both model complexity (degrees of freedom) and differences with the null model, was above the acceptable threshold ( $TLI = 0.94$ ). Finally, the RMSEA, which measures the discrepancy of the model from perfect fit at the population level and considers degrees of freedom, was found to be below the acceptable threshold ( $RMSEA = 0.06$ ).

Based on the dataset of 473 participants, the results of the reliability tests were examined. In the reliability analyses, statistical indicators of the items were considered to evaluate the adequacy of the measurement model. These analyses aimed to empirically support the theoretical framework of the study and to demonstrate the overall structural integrity of the model. The data obtained are presented in Table 9.

According to the findings in Table 9, item-total correlation values were generally above .500, indicating that most items made meaningful contributions to the scale. Removal of items did not lead to significant increases in Cronbach's alpha values. This demonstrates that the scale has a generally high level of reliability ( $\alpha \approx .954\text{--}.956$ ) and that item removal was unnecessary. The results confirmed that the scale possesses an overall reliable structure and is appropriate for measurement purposes.

The results in Table 10 show that the measurement tool under evaluation has high reliability. Both Cronbach's alpha and Guttman's Lambda-2 coefficients exceeded .90, indicating excellent internal consistency, meaning that the scale measures the same construct consistently.

## Conclusion

This study was conducted to develop a reliable and valid scale aimed at measuring secondary school students' self-efficacy in Arabic reading comprehension skills. During the scale development process, a 46-item pool was created by reviewing the relevant literature and obtaining expert opinions. After being evaluated by experts ( $n = 8$ ), this pool was reduced to a 25-item draft form. To test the construct validity of the scale, data were collected from 210 secondary school students, and an exploratory factor analysis (EFA) was performed. According to the EFA results, the KMO value (0.951) and the Bartlett's Test of Sphericity ( $\chi^2 = 2,972.941$ ,  $df = 253$ ,  $p < .000$ ) indicated that the data were suitable for factor analysis. Based on the factor loadings obtained from the EFA, items i6 ("I can understand time expressions (past, present, future, etc.) in an Arabic text I read") and i18 ("I can understand an Arabic dialogue text when I read it") had loadings of 0.291 and 0.289, respectively, falling below the 0.30 threshold, and were therefore excluded from the draft scale. The

scale was thus structured with 23 items under a single dimension. Confirmatory factor analysis (CFA) results revealed that the scale demonstrated good model fit ( $\text{RMSEA} = 0.060$ ;  $\chi^2/\text{df} = 2.59$ ;  $\text{GFI} = .87$ ;  $\text{CFI} = .95$ ). The reliability of the scale was found to be quite high, with a Cronbach's alpha coefficient of .957. Following final revisions, the scale attained its final form with 23 items and a single-factor structure.

When compared with existing measurement tools, such as the *Arabic Speaking Skills Self-Efficacy Scale* by A. Yeşilyurt (2017) and the *Arabic Language Efficacy Questionnaire (ALEQ)* by Mustapha et al. (2013), the current scale occupies a distinct niche. A. Yeşilyurt's (2017) scale focuses on oral language Arabic proficiency without isolating self-efficacy as a central construct. The present study extends this body of work by concentrating exclusively on reading comprehension, thereby offering a more skill-specific and theoretically grounded assessment of self-efficacy.

In terms of practical implications, this scale offers a valuable diagnostic tool for educators and researchers. Teachers in Arabic language programs -particularly in İmam-Hatip high schools where Arabic is taught as a foreign language- can use the scale to identify students' confidence levels in reading comprehension and design instructional strategies to strengthen them. Moreover, the scale's clear theoretical foundation and psychometric robustness make it adaptable to different educational contexts. Future studies are encouraged to validate the instrument in diverse settings, including general high schools, language institutes, and Arabic-speaking countries, to examine its cross-cultural relevance and external validity.

In addition to these findings, the scale's applicability across broader educational contexts should be emphasized. Given that educational systems vary significantly in terms of curriculum structure, pedagogical priorities, and student demographics, the scale has the potential to be adapted for different curricular frameworks with minimal modification. Moreover, its conceptual foundations allow for use in cross-cultural settings, making it a valuable tool for comparative educational research. Future studies could explore the scale's performance in diverse cultural or institutional environments to further establish its generalizability and practical relevance.

While the scale exhibits solid methodological grounding, several limitations warrant attention. The use of convenience sampling and gender imbalance in the EFA stage may limit representativeness. Additionally, the study did not examine predictive validity or test-retest reliability, which should be addressed in future research to evaluate temporal stability and performance-based relationships. Further, although a unidimensional structure was supported statistically, subsequent studies may explore potential subdimensions of Arabic reading self-

efficacy (e.g., inferencing, vocabulary, and cognitive strategy use) to refine the construct.

### **Limitations and Suggestions**

In the scale development literature, it is a widely accepted practice to conduct EFA and CFA on different samples. This approach is important for testing the consistency of the scale across diverse groups. In this study as well, a sample with gender diversity was used at the CFA stage in order to test the performance of the scale across different demographic structures. However, the differences between the EFA and CFA samples may have limited the model fit, and this has been acknowledged as one of the limitations of the study. Future research should replicate the analyses with more homogeneous and balanced samples, thereby allowing model fit and construct validity to be tested more robustly.

The scale presented in this study was originally developed in Turkish. In order to facilitate comprehension for reviewers, editors and international readers, the scale items have been professionally translated into English. This translation is provided solely for the purpose of clarity and accessibility and does not represent an official adaptation study of the instrument into English.

Researchers are encouraged to conduct future adaptation studies of the scale into English and other languages, including comprehensive validity and reliability analyses, to further support its use across different cultural and linguistic contexts. Moreover, the absence of a test-retest reliability analysis is acknowledged as a limitation of the study. Future research should include test-retest procedures to evaluate the temporal stability of the scale more comprehensively.

Based on the results of this study, it is recommended that the "Self-Efficacy Scale for Secondary School Students' Arabic Reading Comprehension Skills" be applied across different demographic groups in order to establish normative values. This would enable the examination of changes in self-efficacy levels by grade level, gender, or school type, and contribute to adapting the scale for use with various age groups. Additionally, developing a digital version of the scale could facilitate the administration process, offering teachers and researchers a more accessible tool. Adapting and translating the scale for use in different countries and regions where Arabic is taught could further allow for cross-cultural comparisons and enhance its international applicability. Employing the scale in studies that examine the Arabic reading comprehension self-efficacy of students from different socio-economic backgrounds could also provide valuable insights into educational equity. Finally, using the scale as a needs assessment tool in teacher education and curriculum development could offer multidimensional contributions to the field.

**Table 11.** Turkish Version of the Self-Efficacy Scale.

1	Arapça bir metin okuduğumda ana fikri genel hatlarıyla anlayabilirim.
2	Resimsiz Arapça bir metin okuduğumda rahatlıkla anlayabilirim.
3	Arapça bir metinde okuduklarımı zihnimde canlandıabilirim.
4	Okuduğum Arapça metinde geçen sayısal ifadeleri diğer ifadelerden ayırt edebilirim.
5	Okuduğum Arapça metinde geçen karşılaşırma ifadelerini ayırt edebilirim.
6	Okuduğum Arapça metinle ilgili sorulara cevap verebilirim.
7	Okuduğum Arapça metinde geçen bağlaç/edatları diğer cümle ögelerinden ayıralabilirim.
8	Okuduğum Arapça metinden hareketle kavram ve görselleri eşleştirebilirim.
9	Arapça metinde verilen ipuçlarından hareketle yeni kelimelerin anımlarını tahmin edebilirim.
10	Okuduğum Arapça metindeki duyu vurgularını anlayabilirim.
11	Benimle ve çevremle ilgili temel kâğıtları açıcı bir konuşma metninde okuyarak anlayabilirim.
12	Arapça kısa mesaj ve duyurulardaki temel düşünceyi okuyup kavrayabilirim.
13	Arapça bir metinde aradığım bilgiyi bulabiliyim.
14	Arapça okuduğum bir metinde sebep-sonuç ilişkisi kurabiliyim.
15	Arapça okuduğum metne yönelik sorular sorabiliyim.
16	Arapça okuduğum metindeki yer/zaman/kişi unsurlarını ayırt edebilirim.
17	Arapça kısa hikâye okuduğumda rahatlıkla anlayabilirim.
18	Arapça yol tarifi okuduğumda anlayabilirim.
19	Arapça bir metin okurken doğru hareke kullanabiliyim.
20	Verilen basit bir konuya ilgili Arapça araştırma yapabiliyim.
21	Verilen basit bir konuya ilgili Arapça kaynak tarayabiliyim.
22	Okuduğum Arapça metinde geçen “gel, git, otur, kalk” gibi yönergeleri anlayabilirim.
23	Basit düzeyde Arapça hazırlanmış grafik, tablo vb. görselleri yorumlayabiliyim.

In conclusion, the “Self-Efficacy Scale for Secondary School Students’ Arabic Reading Comprehension Skills” has been developed as a measurement instrument that meets the criteria of validity and reliability. This scale can be employed to assess self-efficacy perceptions regarding Arabic reading comprehension and may contribute to both instructional practices and research in this area (the final version of the scale is provided in the Table 11).

### Acknowledgments

I extend my sincere gratitude to the field experts for their valuable contributions to this research.

### ORCID iD

Sara Akkuş  <https://orcid.org/0000-0001-8612-3017>

### Ethical Considerations

To conduct the study, approval was first obtained from the Social and Human Sciences Research Ethics Committee of İstanbul University (Date and Reference: February 14, 2024 and 2419875). Subsequently, with the permission granted by the İstanbul Provincial Directorate of National Education (Document No: 99467957, dated March 25, 2024), the scale was administered to students enrolled in İmam-Hatip high schools.

### Consent to Participate

In this study, data were collected from secondary school students under the age of 18. Therefore, a parental consent form

was first sent to the families. Students who received parental approval were then asked to complete the consent form themselves.

### Author Contributions

The research is single-authored.

### Funding

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This study is based on the author’s doctoral dissertation and was previously presented at the International Conference on Language Education and Linguistics (ICLEL 2025) in Kaunas, Lithuania.

### Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Data Availability Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

### References

Aslan, M., Sağıtaş, S., & Özkaynar, K. (2025). Metaverse perception scale development and validation. *SAGE Open*, 15(2), 1–20. <https://doi.org/10.1177/21582440251329126>

Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman and Company.

Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, Article 149. <https://doi.org/10.3389/fpubh.2018.00149>

Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. The Guilford Press.

Council of Europe. (2001). *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge University Press.

Creswell, J. W., & Creswell, J. D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches* (5th ed.). Sage Publications.

Dang, V. T. T. (2024). Foreign language reading anxiety and metacognitive strategies in undergraduates' reading comprehension. *SAGE Open*, 14(2), 1–14. <https://doi.org/10.1177/21582440241279663>

DeVellis, R. F. (2012). *Scale development: Theory and applications*. Sage Publications.

Dörnyei, Z. (2005). *The psychology of the language learner: Individual differences in second language acquisition*. Routledge. <https://doi.org/10.1016/j.system.2015.12.007>

Düzgün, G., & Kirkiç, K. A. (2023). A developmental study of the attitude scale towards teaching Arabic language (AST-TAL): Reliability and validity analysis. *International Journal of Psychology and Educational Studies*, 10(2), 406–421. <https://doi.org/10.52380/ijpes.2023.10.2.1163>

Egel, İ. P. (2009). Learner autonomy in the language classroom: From teacher dependency to learner independency. *Procedia-Social and Behavioral Sciences*, 1(1), 2023–2026. <https://doi.org/10.1016/j.sbspro.2009.01.355>

Graham, S. (2007). Learner strategies and self-efficacy: Making the connection. *Language Learning Journal*, 35(1), 81–93. <https://doi.org/10.1080/09571730701315832>

Güngör, H., & Kan, A. (2020). A Scale development study for the determination of reading self-efficacy of learners of Turkish as a foreign language. *International Journal of Language Academy*, 8(3), 49–62. <http://dx.doi.org/10.29228/ijla.41758>

Hsieh, P. H. P., & Schallert, D. L. (2008). Implications from self-efficacy and attribution theories for an understanding of undergraduates' motivation in a foreign language course. *Contemporary Educational Psychology*, 33(4), 513–532. <https://doi.org/10.1016/j.cedpsych.2008.01.003>

Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>

Kim, D.-H., Wang, C., & Truong, T. N. N. (2024). Psychometric properties of a self-efficacy scale for English language learners in Vietnam. *Language Teaching Research*, 28(4), 1412–1427. <https://doi.org/10.1093/ltr/ctaab090>

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford Publications.

Kurudayioğlu, M., Yazıcı, E., & Göktentürk, T. (2021). Turkish teacher candidates' self-efficacies to use listening strategies scale: A validity and reliability study. *SAGE Open*, 11(2), 1–18.

Leithwood, K. (2007). What we know about educational leadership. In J. M. Burger, C. Webber, & P. Knick (Eds.), *Intelligent leadership* (pp. 41–66). Springer.

Magogwe, J. M., & Oliver, R. (2007). The relationship between language learning strategies, proficiency, age and self-efficacy beliefs: A study of language learners in Botswana. *System*, 35(3), 338–352. <https://doi.org/10.1016/j.system.2007.01.003>

McFall, R. M. (1982). A review and reformulation of the concept of social skills. *Behavioral Assessment*, 4, 1–33. <https://doi.org/10.1007/BF01321377>

Mills, N., Pajares, F., & Herron, C. (2006). A revaluation of the role of anxiety: Self-efficacy, anxiety, and their relation to reading and listening proficiency. *Foreign Language Annals*, 39(2), 276–295. <https://doi.org/10.1111/j.1944-9720.2006.tb02266.x>

Mills, N., Pajares, F., & Herron, C. (2007). Self-efficacy of college intermediate French students: Relation to achievement and motivation. *Language Learning*, 57(3), 417–442. <https://doi.org/10.1111/j.1467-9922.2007.00421.x>

Ministry of National Education (MoNE). (2022). *Arabic language curriculum for Anatolian Imam Hatip high schools (Grades 9 and 10)*. MoNE.

Morgado, F. F. R., Meireles, J. F. F., Neves, C. M., Amaral, A. C. S., & Ferreira, M. E. C. (2018). Scale development: Ten main limitations and recommendations to improve future research practices. *Psicologia: Reflexão e Crítica*, 30(1), 3. <https://doi.org/10.1186/s41155-016-0057-1>

Mustapha, N. H., Mustapha, N. F., Daud, N., & Abd Wahab, M. (2013). Arabic language efficacy questionnaire (ALEQ): Assessing self-efficacy and achievement. *GEMA Online Journal of Language Studies*, 13(1), 155–167.

Pallant, J. (2020). *SPSS survival manual: A step by step guide to data analysis using IBM SPSS* (7th ed.). Routledge.

Raoofi, S., Tan, B. H., & Chan, S. H. (2012). Self-efficacy in second/foreign language learning contexts. *English Language Teaching*, 5(11), 60–73. <http://dx.doi.org/10.5539/elt.v5n11p60>

Sağlam, D., & Arslan, A. (2018). The development of English language skills self-efficacy scale for higher education students. *Psycho-Educational Research Reviews*, 7(2), 1–15.

Shrotryia, V. K., & Dhandha, U. (2019). Content validity of assessment instrument for employee engagement. *SAGE Open*, 9(1), Article 215824401882175. <https://doi.org/10.1177/215824401882175>

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.

Wang, C., & Pape, S. J. (2007). A probe into three Chinese boys' self-efficacy beliefs learning English as a second language. *Journal of Research in Childhood Education*, 21(4), 364–377. <https://doi.org/10.1080/02568540709594601>

Wang, Y., & Sun, P. P. (2024). Development and validation of scales for speaking self-efficacy: Constructs, sources, and relations. *Plos One*, 19(1), 1–19. <https://doi.org/10.1371/journal.pone.0297517>

Warner, L., & Schwarzer, R. (2017). Self-efficacy. In A. Wenzel (Ed.), *The SAGE encyclopedia of abnormal and clinical psychology* (Vol. 7, pp. 3036–3038). SAGE Publications, Inc. <https://doi.org/10.4135/9781483365817.n1210>

Yeşilyurt, A. (2017). Development of self-efficacy perception scale related to Arabic speaking skills . *Electronic Turkish Studies*, 12(6), 881–902. <http://dx.doi.org/10.7827/TurkishStudies.11596>

Yeşilyurt, S., & Çapraz, C. (2018). A roadmap for content validity used in scale development studies. *Erzincan Üniversitesi Eğitim Fakültesi Dergisi*, 20(1), 251–264. <https://doi.org/10.17556/erziefd.297741>

Zimmerman, B. J. (2002). Becoming self-regulated learner. *Theory into Practice*, 41(2), 64–70. [https://doi.org/10.1207/s15430421tip4102\\_](https://doi.org/10.1207/s15430421tip4102_)