**Al-Muyassar: Journal of Arabic Education**
Vol. 4, No. 2, 2025, pp. 561–575
E-ISSN 2829-095X
DOI: https://doi.org/10.31000/6d72tz81
Received: October 2, 2025; Revised: December 4, 2025;
Accepted: December 10, 2025
This work is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License

# Demographic Characteristics and Self-Efficacy in Arabic Language Learning Among Indonesian High School Students: Examining Age, Gender, and Educational Level

# الخصائص الديموغرافية والإحساس بالكفاءة الذاتية في تعلم اللغة العربية لدى طلاب المدارس الثانوية الإندونيسيين: دراسة العمر والجنس والمستوى التعليمي

**Agung Muttaqien[1]*, Mochammad Deddy Soe'aiddy[2], Siti Khumairotuzzahra [3], Fikni Mutiara[4]**
[1,2,3,4]Universitas Djuanda, Indonesia
*Corresponding E-mail: agung.muttaqien@unida.ac.id

**ABSTRACT**

This study examines the relationship between age, gender, and educational level and self-efficacy in learning Arabic among Indonesian high school students. This non-experimental survey involved 1,008 students aged 12–18 years: 332 from junior high schools/madrasahs (SMP/MTs) and 676 from senior high schools/madrasahs (SMA/MA). Self-efficacy was calculated as the average of 16 items with four response categories. Linear regression with robust HC3 covariance estimated the main effects and three two-way interactions, with age centered at 15 years. The model explained 1.4% of the variance in scores (adjusted $R^2 = 0.008$); mean scores were nearly identical across gender (2.93 for males and 2.93 for females) and educational level (2.95 for SMP/MTs and 2.92 for SMA/MA). The age × education level interaction was small (B = 0.058, 95% CI [0.005, 0.110], p = 0.031), but was no longer significant after Bonferroni correction (p = 0.094) or when using standard errors grouped by school label (p = 0.119). Estimates for the 14–16-year-old subgroup showed the same direction but were not precise.The three-factor scale indicated insufficient model fit, so construct validity requires further examination. Additionally, the overlap in age ranges across educational levels limits the separation of the effects of age and education. These findings suggest that demographic characteristics do not sufficiently explain self-efficacy in Arabic language learning. Theoretically, these results underscore the importance of examining learning experiences as a factor more closely linked to the development of self-efficacy. Therefore, teachers should not assume that students' self-efficacy is determined by gender or educational level.

**Keywords**: self-efficacy; Arabic language learning; age; gender; educational level.

**مستخلص البحث**

تتناول هذه الدراسة العلاقة بين العمر والجنس والمستوى التعليمي والكفاءة الذاتية في تعلم اللغة العربية لدى طلاب المدارس الثانوية الإندونيسيين. وشمل هذا الاستطلاع غير التجريبي ١,٠٠٨ طلاب تتراوح أعمارهم بين ١٢ و١٨ عامًا: ٣٣٢ طالبًا من المدارس الإعدادية/المدارس الدينية (SMP/MTs) و٦٧٦ طالبًا من المدارس الثانوية/المدارس الدينية (SMA/MA). وقد تم حساب الثقة بالنفس على أنها متوسط ١٦ بندًا مع أربع فئات للإجابة. واستُخدم الانحدار الخطي مع التباين المشترك HC3 القوي لتقدير التأثيرات الرئيسية وثلاثة تفاعلات ثنائية الاتجاه، مع تحديد العمر ١٥ عامًا كنقطة مركزية. وقد فسّر النموذج ١,٤٪ من التباين في الدرجات ($R^2$ المعدّل = ٠,٠٠٨)؛ وكانت متوسطات الدرجات متطابقة تقريبًا بين الجنسين (٢,٩٣ للذكور و٢,٩٣ للإناث) وبين المستويات التعليمية (٢,٩٥ للمدارس الإعدادية/المدارس الدينية (SMP/MTs) و٢,٩٢ للمدارس الثانوية/المدارس الدينية (SMA/MA)). كان التفاعل بين العمر ومستوى التعليم ضئيلًا (B = 0.058، ٩٥٪ CI [0.005, 0.110]، p = 0.031)، لكنه لم يعد ذو دلالة إحصائية بعد تصحيح بونفيروني (p = 0.094) أو عند استخدام الأخطاء المعيارية المجمعة حسب تصنيف المدرسة (p = 0.119). أظهرت التقديرات الخاصة بالمجموعة الفرعية التي تتراوح أعمارها بين ١٤ و١٦ عامًا نفس الاتجاه، لكنها لم تكن دقيقة. وأشار مقياس العوامل الثلاثة إلى عدم كفاية ملاءمة النموذج، لذا فإن صحة البناء تتطلب مزيدًا من الفحص.

بالإضافة إلى ذلك، فإن التداخل في الفئات العمرية عبر المستويات التعليمية يحد من إمكانية الفصل بين تأثيرات العمر والمستوى التعليمي. تشير هذه النتائج إلى أن الخصائص الديموغرافية لا تفسر بشكل كافٍ الفعالية الذاتية في تعلم اللغة العربية. من الناحية النظرية، تؤكد هذه النتائج على أهمية دراسة تجارب التعلم باعتبارها عاملاً أكثر ارتباطًا بتنمية الفعالية الذاتية. لذلك، يجب ألا يفترض المعلمون أن الفعالية الذاتية للطلاب تتحدد حسب الجنس أو المستوى التعليمي.

**الكلمات المفتاحية**: الكفاءة الذاتية؛ تعلّم اللغة العربية؛ العمر؛ الجنس؛ المرحلة الدراسية

## INTRODUCTION

Self-efficacy refers to an individual's belief in their ability to organize and carry out the actions necessary to achieve a specific goal (Bandura, 1997). Selfefficacy beliefs have been widely examined in academic settings (Pajares, 2003) and they are conceptually distinct from related constructs such as outcome expecctations and selfconcept (Mak & Fancourt, 2020).In learning, this belief relates to task selection, effort, strategy use, and persistence when facing difficulties. Self-efficacy is not the same as actual ability or academic achievement; it reflects an individual's assessment of what they believe they can do. Consequently, students with similar abilities evaluate learning demands differently.

Foreign language learning involves interrelated cognitive, emotional, and social demands. Foreign language classroom anxiety, in particular, has long been described as a distinctive affective response to language learning (Horwitz et al., 2019) Several studies have linked self-efficacy to anxiety as well as reading and listening skills (Mills et al., 2006; Mustapha et al., 2013). (Wang and Sun, 2020) and (Wang et al. 2024) demonstrate that the relationship between self-efficacy and language proficiency may vary depending on the learning context. Meanwhile, Ma and Xu 2025) and Fryer et al. (2025) highlight the importance of learning factors and measurement quality. Although both studies affirm the role of self-efficacy, previous research has shown variations in context and approach, so their findings cannot yet be directly generalized. This underscores the need for further research that takes into account the characteristics of the population and the context of Arabic language learning.

Evidence regarding gender differences in foreign language self-efficacy remains mixed. Wang et al. (2024), who studied 606 Chinese college students learning English as a foreign language, examined the potential role of gender in the relationship between perceived classroom goal structure, self-efficacy, and engagement. Ma and Xu (2025) reported that the role of self-efficacy in the relationship between critical thinking dispositions and foreign language

achievement differs by gender, with a stronger mediating effect among male participants. These studies tested different models from the present study; their relevance lies in the need to explicitly examine gender-based patterns, rather than assuming that one group consistently has higher self-efficacy. At a boarder level, a meta analysis of gender differences in academic self-efficacy reported only a small overall difference favoring males, with variation ocrosss domains (Huang, 2013)

In Arabic language education, Mustapha et al. (2013) developed the Arabic Language Efficacy Questionnaire (ALEQ) by involving 234 Arabic language majors to examine self-efficacy and language achievement. However, the available evidence largely involves college students or specific language skills and does not show how age, gender, and educational level are collectively related to general self-efficacy scores among Indonesian high school students.

Theoretically, age, gender, and educational level do not directly determine self-efficacy. According to social-cognitive theory, self-efficacy develops through mastery experiences, vicarious experiences, social persuasion, as well as physiological and emotional conditions (Bandura, 1997; Schunk & Pajares, 2002). Whit this framework, self-efficacy operates through reciprocal interactions among personal, behavioral, and environmental influences. This study examines all three demographic characteristics simultaneously to understand their relationship with self-efficacy in Arabic language learning. and mastery experience is typically the most influential source, although the strength of each source varies of context such as gender, ability, and academic domain(Usher,E. L., & Pajares, 2008). Its main contribution lies in testing the separability of the effects of age and educational level, given that the two may be interrelated and difficult to distinguish when the age ranges across groups do not overlap. Thus, this study not only identifies demographic associations with self-efficacy but also highlights the interpretive limitations of these associations in the context of Arabic language education.

This study addresses the limited evidence regarding the combined patterns of age, gender, and educational level in relation to self-efficacy in Arabic language learning among high school students, while also testing the robustness of the findings across various model specifications. Given the diverse literature, this study is exploratory in nature and does not establish directional hypotheses. The research questions are: How are age, gender, and educational level related to self-efficacy scores? Do the relationships between age and self-efficacy differ by gender or

educational level? Does gender interact with educational level? And do the interaction findings hold after controlling for multiple testing, clustered inferences, and sensitivity analysis? Therefore, this study examines the relationships and interactions among these characteristics regarding self-efficacy in learning Arabic among junior high school (SMP/MTs) and senior high school (SMA/MA) students. All findings are interpreted as associations, not causal effects.

## RESEARCH METHOD

This study employed a non-experimental quantitative survey design. Data collection was conducted via an online questionnaire using Google Forms, which was distributed with the assistance of Arabic language teachers at each school. The respondents were junior high school (SMP/MTs) and senior high school (SMA/MA) students currently taking Arabic language classes in 26 provinces across Indonesia. Respondent recruitment was based on accessibility through the network of Arabic language teachers; therefore, this study employed a non-probability sampling approach rather than random sampling. The provinces with the highest number of participants were Banten (n = 235), West Java (n = 175), Southeast Sulawesi (n = 118), West Sumatra (n = 110), and Central Kalimantan (n = 80). After data screening, 1,008 participant records were retained for analysis: 332 from junior high schools/MTs (32.9%) and 676 from senior high schools/MA (67.1%); 410 males (40.7%) and 598 females (59.3%). Participants' ages ranged from 12 to 18 years (M = 15.59, SD = 1.80), and their recorded grade levels ranged from 7th to 12th grade. The available analysis files did not adequately document the methods and periods of data collection, the procedures for selecting and recruiting schools/classes, details of ethical approval, or the number of independently verified schools. Therefore, this sample should not be considered a national probability sample. Normalized school labels yielded 347 unique labels; however, these labels have not been verified as representing 347 distinct schools. Unusual grade-level records were retained as recorded; without source verification, these records cannot be definitively classified as data entry errors or valid cases. Original documentation must be used to supplement details regarding participation, consent for minors, and confidentiality prior to submission.

Self-efficacy was measured using 16 items (ED_01–ED_16) with a four-category response scale scored from 1 to 4. This instrument was designed to cover three aspects: ability, learning strategies, and perseverance. However, the source or

basis for the instrument's adaptation, verbal response labels, translation process, expert review, and pilot testing were not documented in the available analysis files. Eight negatively worded items (ED3, ED4, ED7, ED8, ED11, ED12, ED15, and ED16) were reverse-scored using the formula 5 minus the original score. The composite score ED_MEAN is the average of these 16 items; a higher score indicates a higher level of self-confidence (M = 2.93, SD = 0.39). Internal consistency in the analysis dataset was Cronbach's α = 0.879. A three-factor confirmatory factor analysis (CFA) indicated an inadequate fit (CFI = 0.937, TLI = 0.926, RMSEA = 0.141, SRMR = 0.107; robust CFI = 0.708, robust TLI = 0.654, robust RMSEA = 0.154), with inter-factor correlations ranging from 0.910 to 0.954. these values fall short of commonly cited benchmarks of approximately. 95 for CFI and TLI, 06 for RMSEA, and 08 for SRMR (Clarifications et al., 2025). Results for the one-factor model and the method factor for negatively worded items are not available for comparison. Measurement invariance based on gender and educational level was not tested. Therefore, ED_MEAN was treated as an observed composite score, not as a validated latent construct score.

The main analysis involved a regression of the ED_MEAN variable on age centered at 15 years, gender (reference category: male), educational level (reference category: junior high school/MTs), and three two-way interactions: age × gender, age × educational level, and gender × educational level. Coefficient estimates were calculated using robust HC3 covariance and 95% confidence intervals. Since three interactions were tested, the p-value for each interaction was adjusted using the Bonferroni method (p × 3). A combined test of the three interactions, as well as the change in $R^2$ ($\Delta R^2$) attributable to the interaction blocks, was also reported. The reporting of interactions followed recommendations for presenting direction, effect size, estimation precision, and visualization (Murphy & Aguinis, 2022; Pek & Flora, 2018).

The robustness of the age × education level interaction was tested exploratively by using standard errors grouped by normalized school labels, restricting the sample to ages 14–16, replacing age with numerical grade values, and adding the age-squared term and its interaction. There were 347 normalized school labels; however, school identities and cluster size distributions were not verified; therefore, the clustered results were treated as sensitivity analyses rather than definitive cluster-based conclusions. The 14–16-year-old age subgroup included 461 students. All sensitivity analyses are exploratory and were not adjusted separately

for multiple testing; interpretations are based on the main model and its multiple-testing corrections. Observed means based on the age × education level interaction are presented with 95% confidence intervals only for cells with n ≥ 10. The software and its version are not documented in the available files and must be obtained from , the original analysis notes.

## RESULT AND DISCUSSION

### *Descriptive Statistics*

Table 1 presents the means, standard deviations, and sample sizes by gender and educational level. The marginal ED_MEAN scores were 2.934 (SD = 0.381, n = 410) for males and 2.929 (SD = 0.397, n = 598) for females; the mean was 2.954 (SD = 0.387, n = 332) for junior high school/MTs students and 2.920 (SD = 0.391, n = 676) for senior high school/MA students. These marginal differences were very small (≤ 0.04 points on a 1–4 scale).

**Table 1. Descriptive Statistics of Self-Efficacy by Gender and Educational Level**

| Gender | Educational level | *n* | *M* | *SD* |
|---|---|---|---|---|
| Male | SMP/MTs | 143 | 2.913 | 0.357 |
| Male | SMA/MA | 267 | 2.945 | 0.393 |
| Female | SMP/MTs | 189 | 2.986 | 0.407 |
| Female | SMA/MA | 409 | 2.903 | 0.390 |

**Source: research data. Scale range 1–4; higher scores indicate higher self-efficacy.**

### *Age Distribution and Group Averages*

Age distribution differed significantly by educational level: junior high school (SMP/MTs) students were mostly 12–14 years old, while senior high school (SMA/MA) students were mostly 16–18 years old. At age 14, there were 115 junior high school (SMP/MTs) students and 3 senior high school (SMA/MA) students; at age 16, there were 2 junior high school (SMP/MTs) students and 207 senior high school (SMA/MA) students (Table 2). Cells with very small sample sizes also occurred among high school (SMA/MA) students aged 12 and 13 (n = 1 each) and junior high school (SMP/MTs) students aged 16 and 17 (n = 2 each). Source data for these uncommon age–education level combinations have not been independently verified; therefore, these small cells should be interpreted with caution and should not be used to support strong conclusions.

**Table 2. Mean Self-Efficacy Scores by Age and Educational Level**

| Age | Educational level | *n* | *M* | *SD* | 95% CI mean |
|---|---|---|---|---|---|
| 12 | SMP/MTs | 52 | 3.029 | 0.423 | [2.911, 3.147] |
| 12 | SMA/MA | 1 | 2.312 | — | — |
| 13 | SMP/MTs | 121 | 2.922 | 0.407 | [2.849, 2.995] |
| 13 | SMA/MA | 1 | 3.312 | — | — |
| 14 | SMP/MTs | 115 | 2.974 | 0.360 | [2.908, 3.041] |
| 14 | SMA/MA | 3 | 3.062 | 0.348 | — |
| 15 | SMP/MTs | 40 | 2.925 | 0.349 | [2.813, 3.037] |
| 15 | SMA/MA | 94 | 2.910 | 0.361 | [2.836, 2.983] |
| 16 | SMP/MTs | 2 | 2.656 | 0.044 | — |
| 16 | SMA/MA | 207 | 2.885 | 0.371 | [2.834, 2.936] |
| 17 | SMP/MTs | 2 | 2.719 | 0.575 | — |
| 17 | SMA/MA | 203 | 2.872 | 0.396 | [2.817, 2.927] |
| 18 | SMA/MA | 167 | 3.024 | 0.410 | [2.962, 3.087] |

**Source: research data. The 95% confidence interval is shown only for cells with n ≥ 10; mean values for smaller cells should be interpreted with great caution.**

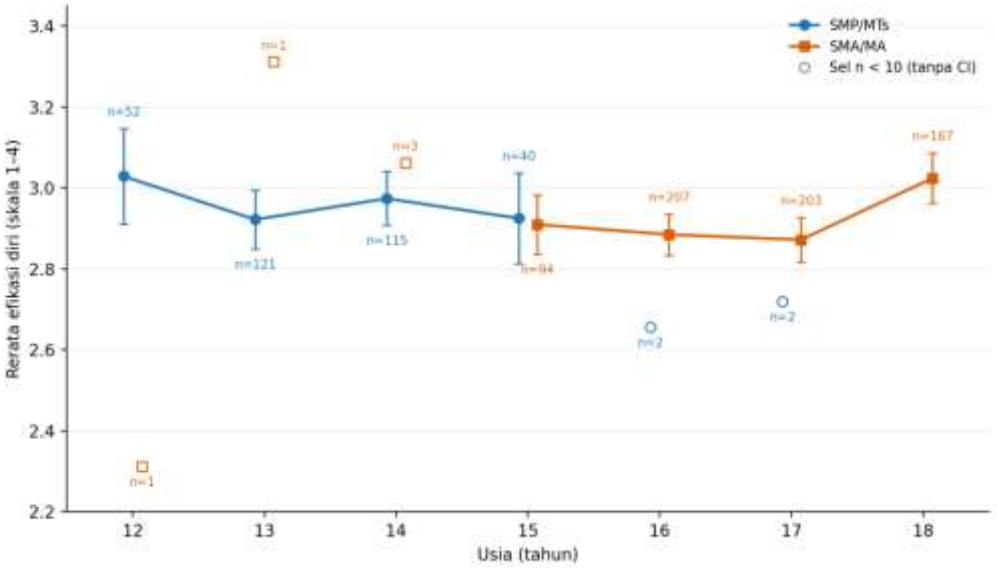



**Figure 1. Observed Self-Efficacy Means by Age and Educational Level (error bars: 95% CI; open markers: cells with n < 10, without CI or connecting lines)**

**Source: research data**

### *Primary Model and Interactions*

The results of the regression analysis show that the overall research model explains approximately 1.4% of the variation in students' self-efficacy scores ($R^2$ = 0.0137; adjusted $R^2$ = 0.0078), with the results of the conventional OLS test being significant, F(6, 1001) = 2.32, p = 0.031. Adding the interaction block to the main effects model increased the $R^2$ value by 0.0096, F(3, 1001) = 3.26, p = 0.021. The results of the combined test using the HC3 standard error also indicated that the interaction

term was significant, $F(3, 1000) = 3.07$, $p = 0.027$. Nevertheless, the low $R^2$ value suggests that the demographic characteristics included in the model have only limited explanatory power regarding the variation in students' self-efficacy in Arabic language learning.

These findings suggest that self-efficacy in Arabic language learning is likely not only related to demographic characteristics but is also influenced by various learning experiences and conditions not accounted for in this study's model. This is consistent with Bandura's self-efficacy theory, which explains that an individual's belief in their own abilities is formed through four main sources: *mastery experiences*, *vicarious experiences*, *social persuasion*, and *physiological and emotional states*. In the context of Arabic language learning, experiences of success in understanding the material, completing exercises, and using Arabic can serve as important sources in shaping students' beliefs about their learning abilities.

In addition to individual experiences, factors within the learning environment also have the potential to contribute to the development of self-efficacy. Teacher characteristics, learning strategies and methods, the quality of interactions between teachers and students, and the academic climate at school can provide different learning experiences for each student. Family support, motivation to learn, experiences receiving feedback, and opportunities to use Arabic can also be relevant factors to consider. These factors have the potential to explain some of the variation in self-efficacy not accounted for in this study's model.

Thus, the low $R^2$ value does not necessarily indicate that demographic characteristics are irrelevant to the development of self-efficacy. Rather, these findings suggest that the contribution of demographic characteristics in explaining self-efficacy is relatively limited when analyzed using the model employed. Therefore, future research should consider incorporating variables related to learning experiences, teacher characteristics, teaching approaches, the school environment, and family support to gain a more comprehensive understanding of the factors associated with students' self-efficacy in Arabic language learning.

The degrees of freedom in the denominator for the robustness test follow the results of the available HC3 analysis. The coefficients are presented in Table 3. Since this model includes interactions, the main effects are conditional on the reference category and age 15; they do not represent overall differences across the entire

sample.

**Table 3. Regression Coefficients for the Model with Two-Way Interactions**

| Predictor | *B* | *SE* (HC3) | 95% CI B | *p* |
|---|---|---|---|---|
| Intercept (age 15; male; SMP/MTs) | 2.879 | 0.048 | [2.784, 2.974] | < 0.001 |
| Female | 0.076 | 0.055 | [−0.033, 0.184] | 0.172 |
| SMA/MA | 0.010 | 0.071 | [−0.129, 0.149] | 0.885 |
| Female × SMA/MA | −0.125 | 0.091 | [−0.304, 0.054] | 0.172 |
| Age (centred at 15 years) | −0.024 | 0.026 | [−0.075, 0.027] | 0,358 |
| Age × female | 0.005 | 0.025 | [−0.045, 0.054] | 0,856 |
| Age × SMA/MA | 0.058 | 0.027 | [0.005, 0.110] | 0,031 |

**Source: research data. Reference category: males and junior high school/MTs. SE = robust HC3 standard error. P-values in the table are not adjusted for repeated testing.**

Of the three interactions, only the age × education level interaction had a p-value < 0.05 before correction (B = 0.058, 95% CI [0.005, 0.110], p = 0.031). After Bonferroni correction for these three interactions, p = 0.094; with standard errors clustered by school label, p = 0.119 (Table 4). The age × gender interaction (clustered p = 0.851) and the gender × educational level interaction (clustered p = 0.165) were not statistically significant in either of the two inference approaches.

Positive interaction coefficients indicate that the age slope is estimated to be greater for SMA/MA than for SMP/MTs. The age slope for junior high schools (SMP/MTs) is −0.024 points per year (95% CI [−0.075, 0.027], p = 0.358), while the estimated slope for senior high schools (SMA/MA) is approximately 0.034 points per year. Confidence intervals and p-values for the high school/MA slope as a linear combination were not reported because the required covariance output was not available in the provided analysis materials. A difference in slope of 0.058 points per year on a scale of 1–4 is considered small.

**Table 4. Age × Education Level Interaction Across All Analyses**

| Predictor | *B* | *SE* (HC3) | 95% CI B | *p* |
|---|---|---|---|---|
| Intercept (age 15; male; SMP/MTs) | 2.879 | 0.048 | [2.784, 2.974] | < 0.001 |
| Female | 0.076 | 0.055 | [−0.033, 0.184] | 0.172 |
| SMA/MA | 0.010 | 0.071 | [−0.129, 0.149] | 0.885 |
| Female × SMA/MA | −0.125 | 0.091 | [−0.304, 0.054] | 0.172 |
| Age (centred at 15 years) | −0.024 | 0.026 | [−0.075, 0.027] | 0,358 |
| Age × female | 0.005 | 0.025 | [−0.045, 0.054] | 0,856 |
| Age × SMA/MA | 0.058 | 0.027 | [0.005, 0.110] | 0,031 |

**Source: study data. B represents the difference in the age slope for SMA/MA relative to SMP/MTs. Subgroup and grade analyses are exploratory; grade is strongly associated with educational level, so the grade model is not an independent replication.**

*Sensitivity and Exploratory Analyses*

In the 14–16-year-old subgroup (n = 461; 304 high school students and 157 middle school students), the age × education level interaction was B = 0.036 (95% CI [−0.099, 0.170], p = 0.605). This estimate has the same direction as in the overall sample, but the interval is much wider and includes both zero and the estimate from the overall sample (0.058). This subgroup also remains unbalanced: there were only 3 high school students aged 14, and only 2 middle school students aged 16.

When numerical scores were used instead of age, the class × education level interaction was B = 0.064 (95% CI [0.003, 0.125], p = 0.040; n = 1,008), consistent with the possibility that score trends differ by education level. A combined HC3-robust test of three quadratic terms (age², age² × gender, and age² × educational level) yielded F(3, 998) = 2.86, p = 0.036, suggesting that a linear specification of age may be inadequate. However, small cell sizes and imbalances in the age–education level cross-tabulation preclude definitive interpretations regarding the shape of group-specific curves.

**Discussion**

The model, which includes age, gender, educational level, and two-way interactions among these variables, explains only 1.4% of the variance in self-efficacy. Marginal means are nearly identical between males and females as well as between junior high school/MT and senior high school/MA (differences ≤ 0.04 points on a 1–4 scale). This is consistent with the view that self-efficacy is related to proximal learning experiences—such as experiences of mastery, feedback, social modeling, and emotional states—rather than merely demographic characteristics (Bandura, 1997; Schunk & Pajares, 2002). Since these experiences were not measured in this study, this explanation serves as an interpretive framework rather than a mechanism substantiated by the data.

Research in language education indicates that students' self-efficacy is influenced by learning experiences and psychological conditions rather than demographic characteristics alone. (Calafato, 2023) in a study of Arabic language learners, found that motivation, learning strategies, and self-efficacy were interconnected and could fluctuate in response to students' learning experiences. Similarly, (Zhang, 2024) demonstrated that self-regulation strategies contributed to improvements in language learning motivation and self-efficacy. In the context of Arabic learning, (Mei et al., 2023)(Noor, 2026) found that anxiety negatively affected

students' classroom performance, whereas self-efficacy had a positive effect. More recently, (Maharani & Zalni, 2026) highlighted the roles of anxiety, motivation, self-efficacy, and classroom interactions in Arabic language learning. These findings provide a possible explanation for the limited contribution of demographic characteristics observed in the present study (Wei et al., 2026). However, since learning experiences, motivation, and emotional conditions were not directly measured, their influence should be considered a possible interpretation rather than an established explanation of the findings.(Mei et al., 2023)

The evidence that educational level moderates the relationship between age and self-efficacy is not strong. The interaction block of these three variables adds $\Delta R^2$ of 0.0096 and is significant as a block ($p = 0.021$ with OLS; $p = 0.027$ with HC3), but the age × education level interaction coefficient is no longer significant after Bonferroni correction ($p = 0.094$) or with clustered standard errors ($p = 0.119$). In the 14–16-year-old age subgroup, the estimates had the same direction but were not precise, and their confidence intervals encompassed the estimates for the overall sample; therefore, this subgroup neither confirmed nor refuted the presence of an interaction. Limited overlap in the age distribution made comparisons of slopes dependent on very small cells. Therefore, this interaction should be treated as an exploratory finding rather than a confirmation. The class model ($p = 0.040$) and quadratic test ($p = 0.036$) were also exploratory, unadjusted, and do not constitute independent evidence.

There is insufficient evidence that the age–self-efficacy relationship differs by gender or that gender interacts with educational level. Nonsignificant results do not prove equality across groups; rather, they indicate that the data do not provide adequate support for the tested differences in slopes. This interpretation aligns with the caution that gender patterns in foreign language learning are not uniform (Ma & Xu, 2025; Wang et al., 2024), although those studies examined different models and populations. Existing studies on self-efficacy in Arabic also involved different student populations and scales (Mustapha et al., 2013), so their results cannot be directly compared with those for high school students in this study.

The quality of the measurement limits interpretation. Cronbach's $\alpha = 0.879$ indicates internal consistency, but the three-factor CFA did not show adequate fit, and the correlations between factors were very high (0.910–0.954), suggesting that the three proposed dimensions are not empirically distinct. Measurement invariance

across gender and educational level was not tested. Therefore, ED_MEAN was treated as an observed composite score, and no conclusions were drawn regarding differences in latent constructs across groups (see Fryer et al., 2025, regarding the importance of measurement quality).

Limitations of this study include its observational design, which precludes causal inferences; a sample that cannot be assumed to be representative of the national population; and an age–education level imbalance, which limits the ability to disentangle age effects from education level effects. School labels were not fully verified, and class identifiers were not available, so groupings could only be estimated. School type (madrasah versus public school) is associated with educational level and was not included as a covariate. Furthermore, the source of the instruments and the validation process were not adequately documented, the factor structure was not supported by CFA results, and measurement invariance was not tested.

Future research should use a sample with an overlapping age range across all educational levels, verified school and class identifiers for hierarchical analysis, validated instruments with measurement invariance testing, and direct measures of sources of self-efficacy (e.g., experiences of mastery, teacher feedback, and academic anxiety) that are theoretically more closely related to its development.

## CONCLUSION

Among 1,008 junior high school (SMP/MTs) and senior high school (SMA/MA) **s**tudents, age, gender, educational level, and two-way interactions among these variables explained approximately 1.4% of the variance in self-efficacy for learning Arabic, with relatively similar mean scores across all groups. The age × educational level interaction (B = 0.058, 95% CI [0.005, 0.110]) was not significant after correction for multiple testing (p = 0.094) or clustered inference (p = 0.119), nor was it confirmed in the 14–16-year-old subgroup.

Overall, these findings suggest that demographic characteristics have limited explanatory power in understanding self-efficacy for learning Arabic. However, this interpretation must account for limitations in measurement validity, age–educational level imbalances, and unverified grouping structures. Therefore, teachers should not assume differences in self-efficacy based on gender or educational level but should instead consider the individual characteristics and learning experiences of each

student. Future research should prioritize the use of validated instruments, overlapping age ranges, and verified school identities. Additionally, direct measurements of mastery experiences, vicarious experiences, social persuasion, and physiological and emotional states are necessary to gain a deeper understanding of the development of self-efficacy in Arabic language learning.

## REFERENCES


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

Calafato, R. (2023). Charting the motivation, self-efficacy beliefs, language learning strategies, and achievement of multilingual university students learning Arabic as a foreign language. Asian-Pacific Journal of Second and Foreign Language Education, 5. https://doi.org/10.1186/s40862-023-00194-5

Clarifications, S., Many, F. I., Indices, A. F., Ii, T., West, S., Criterion, I., Criterion, B. I., & Cross-, E. (2025). χ 2 + 2(. 1–5.

Fryer, L. K., Li, C., Guo, Z., Liang, X., & Zhong, Y. (2025). The role of self-efficacy in second language learning during formal education: A systematic review, critical evaluation of previous research, and future directions for research and practice. Studies in Second Language Learning and Teaching, 15(2), 251–278. https://doi.org/10.14746/ssllt.48234

Horwitz, E. K., Horwitz, M. B., Cope, J., Horwitz, E. K., Horwitz, M. B., & Cope, J. (2019). Foreign language classroom anxiety. 70(2), 125–132.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis.

Ma, L., & Xu, L. (2025). Critical thinking attitudes and foreign language learning: A moderated mediation analysis of self-confidence and gender. International Journal of Educational Research, 133, 102740. https://doi.org/10.1016/j.ijer.2025.102740

Maharani, N., & Zalni, N. (2026). A systematic review of psychological and sociological factors in. 11(1), 253–265. https://doi.org/10.18860/abj.v11i1.39913

Mak, H. W., & Fancourt, D. (2020). Longitudinal associations between reading for pleasure and child maladjustment: Results from a propensity score matching analysis. Social Science & Medicine, 253, 112971. https://doi.org/10.1016/j.socscimed.2020.112971

Mei, S. Y., Shittu, M. A., & Ju, S. Y. (2023). Effect of anxiety and self-efficacy on class performance in Arabic language online class. World Journal of English Language, 13(5), 269–276. https://doi.org/10.5430/wjel.v13n5p269

Mills, N., Pajares, F., & Herron, C. (2006). A re-examination of the role of anxiety: Self-efficacy, anxiety, and their relationship to reading and listening proficiency. Journal of Foreign Languages, 39(2), 276–295.

https://doi.org/10.1111/j.1944-9720.2006.tb02266.x

Murphy, K. R., & Aguinis, H. (2022). Reporting interaction effects: Visualization, effect sizes, and interpretation. Journal of Management, 48(8), 2159–2166. https://doi.org/10.1177/01492063221088516

Mustapha, N. H., Mustapha, N. F., Daud, N., & Wahab, M. A. (2013). The Arabic Language Efficacy Questionnaire (ALEQ): Assessing self-efficacy and achievement. GEMA Online Journal of Language Studies, 13(1), 155–167. https://ejournal.ukm.my/gema/article/view/2232

Noor, F. (2026). Foreign language anxiety in Arabic speaking and writing: A systematic review of correlates and interventions, 2(1).

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. Reading & Writing Quarterly.

Pek, J., & Flora, D. B. (2018). Reporting effect sizes in original psychology research: A discussion and guide. Psychological Methods, 23(2), 208–225. https://doi.org/10.1037/met0000126

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. In A. Wigfield & J. S. Eccles (Eds.), The development of achievement motivation (pp. 15–31). Academic Press. https://doi.org/10.1016/B978-012750053-9/50003-6

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions.

Wang, C., & Sun, T. (2020). The relationship between self-confidence and language proficiency: A meta-analysis. System, 95, 102366. https://doi.org/10.1016/j.system.2020.102366

Wang, H., Patterson, M. M., & Long, H. (2024). Student engagement in foreign language learning: Its relationship with classroom goal structure, self-efficacy, and gender. Frontiers in Education, 9, 1416095. https://doi.org/10.3389/feduc.2024.1416095

Wei, N., Seet, K., Kadir, M. A. A., Suárez, L., & Dillon, D. (2026). Second language anxiety and enjoyment among primary school-aged students: A systematic review, 2.

Zhang, T. (2024). Effects of self-regulation strategies on EFL learners' language learning motivation, willingness to communication, self-efficacy, and creativity. 1–13.