



## **Development of an Augmented Reality-Based Vocabulary Builder on Recipe Procedure Text to Improve the Vocabulary Mastery of Vocational School EFL Students**

**Raihan Anindra Asri, Hermi Zaswita\***

English Language Education Study Program, Department of English Language and Literature, Faculty of Languages and Arts, Universitas Negeri Padang, Indonesia.

\*Corresponding Author. Email: [hermizaswita@fbs.unp.ac.id](mailto:hermizaswita@fbs.unp.ac.id)

**Abstract:** This study aims to develop an augmented reality (AR)-based Vocabulary Builder for eleventh-grade Culinary students at SMK Negeri 9 Padang that presents culinary vocabulary as three-dimensional objects with definitions and links each word to recipe use through exercises and a recipe-completion quiz. A research and development (R&D) design was used, following the 4D model (Define, Design, Develop, and Disseminate). Validity was judged by three experts and analysed with Aiken's V, practicality was judged by three English teachers and 20 students, and effectiveness was tested with a one-group pretest-posttest design involving 30 Grade XI Culinary students in six meetings, using a 40-item vocabulary test. The product was highly valid ( $V = .91$ ), and both teachers (91.0%) and students (89.0%) rated it very practical. Mean test scores rose from 72.33 ( $SD = 18.30$ ) to 87.93 ( $SD = 10.42$ ). Because the scores were not normally distributed, a Wilcoxon signed-rank test was used,  $Z = 4.77$ ,  $p < .001$ ,  $r = .87$ , and the normalized gain was medium ( $g = .56$ ); 29 of the 30 students improved. The findings suggest that the AR Vocabulary Builder is a valid, practical, and promising medium for vocational EFL vocabulary learning, although the one-group design and a ceiling effect at posttest limit causal claims. Further studies should add a control group and a delayed posttest.

### **Article History**

Received: 02-08-2026

Revised: 17-09-2026

Accepted: 04-10-2026

Published: 20-10-2026

### **Key Words:**

Augmented Reality;  
Vocabulary Builder;  
Vocational School;  
Recipe Procedure Text.

**How to Cite:** Asri, R. A., & Zaswita, H. (2026). Development of an Augmented Reality-Based Vocabulary Builder on Recipe Procedure Text to Improve the Vocabulary Mastery of Vocational School EFL Students. *Jurnal Paedagogy*, 13(4), 1755-1766. <https://doi.org/10.33394/jp.v13i4.22724>



<https://doi.org/10.33394/jp.v13i4.22724>

This is an open-access article under the [CC-BY-SA License](#).



## **Introduction**

Vocabulary is a foundation of English proficiency. Learners need a large stock of words to understand what they read and hear (Nation, 2006) and to express their ideas in speech and writing; as Wilkins (1972) observed, “without vocabulary nothing can be conveyed.” In Indonesian EFL contexts, vocabulary knowledge is closely related to learners’ overall English competence (Hartini & Ardini, 2024). Knowing a word involves knowing its structure and meaning, and the same applies to formulaic sequences such as idioms and collocations (Vu & Peters, 2021). Nation (2022) and Webb and Nation (2017) therefore recommend a balanced programme built on four strands: meaning-focused input, meaning-focused output, language-focused learning, and fluency development. The effectiveness of L2 vocabulary instruction has also been synthesised meta-analytically (Yousefi & Biria, 2018); what remains open is which medium suits a particular group of learners.

Vocabulary is a foundation of English proficiency. Learners need a large stock of words to understand what they read and hear (Nation, 2006) and to express their ideas in speech and writing; as Wilkins (1972) observed, “without vocabulary nothing can be conveyed.” In Indonesian EFL contexts, vocabulary knowledge is closely related to learners’



overall English competence (Hartini & Ardini, 2024). Which medium best suits a particular group of learners, however, remains open.

Students in vocational high schools (Sekolah Menengah Kejuruan, SMK) face a double demand: they must learn general English vocabulary and the specialised vocabulary of their chosen field. In the Culinary program, the eleventh-grade English syllabus includes procedure texts, which explain how something is done through a sequence of steps (Gerot & Wignell, 2015; Nabujing & Utari, 2023). To work with a recipe, students must identify its goal, its ingredients and equipment, and its ordered cooking actions. Ingredient names, kitchen utensils, cooking verbs, and measurements are therefore the very words that carry a recipe's meaning. Maula (2021) developed English learning materials for a culinary arts program with the four-D model, which shows that vocational English materials can be developed systematically for this field.

The problem is evident at SMK Negeri 9 Padang. Preliminary observations, interviews with students, and school records showed that eleventh-grade students had a limited vocabulary repertoire, had difficulty recalling words they had studied, and were reluctant to speak English because they lacked the words. Daily quiz averages on procedure text ranged from 50 to 70, below the minimum passing score (Kriteria Ketuntasan Minimal, KKM) of 75, and Table 1 shows that only 38.2% to 45.7% of the students in the three Grade XI Culinary classes passed the unit in the January–June 2026 semester. Teachers relied on teaching modules and a whiteboard, no interactive media were in use, and students said that recipe vocabulary was hard to understand. The school also has only one kitchen, so opportunities for students to practise with real kitchen objects are limited.

**Table 1. Students' Results in the Procedure Text Unit, January–June 2026 Semester**

| Class         | Total students | Passed, n (%)    | Not passed, n (%) |
|---------------|----------------|------------------|-------------------|
| XI Culinary 3 | 34             | 13 (38.2)        | 21 (61.8)         |
| XI Culinary 4 | 35             | 15 (42.9)        | 20 (57.1)         |
| XI Culinary 5 | 35             | 16 (45.7)        | 19 (54.3)         |
| <b>Total</b>  | <b>104</b>     | <b>44 (42.3)</b> | <b>60 (57.7)</b>  |

*Note.* Data were taken from the SMK Negeri 9 Padang school database (2026). The minimum passing score (KKM) was 75. Totals and their percentages were calculated by the authors.

Augmented reality (AR) is a promising response to these problems. AR combines real and virtual objects in a real environment, works interactively in real time, and aligns virtual and real objects with each other (Azuma et al., 2001; Marrahí-Gómez & Belda-Medina, 2023). For vocabulary learning, this means that a word can be attached to a three-dimensional object that students inspect through the camera of their own device, in effect a virtual kitchen that compensates for the school's single kitchen.

In vocational English, recipe words are not abstract labels: each names an object, a manual action, or a measure that learners will meet again in the kitchen. When such words are learned only from modules and a whiteboard, they stay detached from what they refer to. The AR Vocabulary Builder bridges this gap in three steps. A word is first attached to a three-dimensional model that students inspect through their device camera, together with its name, definition, and pronunciation; it is then practised in a multiple-choice exercise; and it is finally placed back into a real procedure through a recipe-completion quiz built around one recipe, pizza margherita. Learners thus move from word to object to recipe step.

Empirical studies have examined AR for vocabulary learning among young learners (Binhomran & Altalhab, 2024; Korosidou, 2024; Lai & Chang, 2021), in general EFL classes (Belda-Medina & Marrahi-Gomez, 2023; Khan et al., 2023; Sadikin & Martyani, 2020; Tsai,



2020), and in university ESP courses (Amin, 2025). Several studies report advantages of AR over conventional or flashcard-based instruction in vocabulary performance and motivation (Lai & Chang, 2021; Sadikin & Martyani, 2020; Tsai, 2020). The evidence is not uniformly positive, however. A meta-analysis of six studies of AR and English vocabulary found an effect that favoured AR but did not reach statistical significance (Direkci & Aydemir, 2026), and AR visualisations can help students learn some kinds of content while hindering others (Radu & Schneider, 2023). Design quality therefore matters, and cognitive load and multimedia design principles need careful handling in AR environments (Buchner et al., 2022; Krüger & Bodemer, 2022).

Two gaps remain. First, many of these studies involved young learners or general education, and empirical evidence on AR vocabulary media built around a vocational field, here culinary procedure texts, in Indonesian vocational schools remains limited. Second, AR products developed in Indonesia are often reported as prototypes tested for feasibility or usability (Agata et al., 2021; Hadi, 2025; Kosniawati et al., 2026), without bringing expert validity, classroom practicality, and effectiveness together in one study. A development study that follows a recognised model and reports all three quality criteria (Plomp & Nieveen, 2013) can address both gaps.

Recipe-procedure vocabulary is particularly suited to three-dimensional, spatial presentation for three reasons. First, much of it names objects, such as equipment and utensils, whose shape, size, and parts must be seen from several angles to be understood. Second, many procedural verbs, such as fold, spread, or place on, describe actions with spatial meaning that a flat picture or a line of text conveys poorly. Third, a recipe is a sequence of actions on physical objects, so learners must link each word to a referent and to its place in the sequence. A model that learners can rotate, enlarge, and hear named supports these links, in line with multimedia learning, in which words and pictures presented together deepen learning when working-memory limits are respected (Mayer, 2020, 2024), and with situated vocabulary learning (Santos et al., 2016; Larchen Costuchen et al., 2021).

This study therefore aims to develop an AR-based Vocabulary Builder for recipe procedure texts for Grade XI Culinary students and to test its validity, practicality, and effectiveness. It contributes in three ways. Practically, it offers a ready-to-use AR vocabulary tool for a vocational school with a single kitchen, with quality evidence from experts, teachers, and students. Theoretically, it extends research on AR-assisted vocabulary learning, which has mostly involved young learners or general EFL classes, to technology-enhanced English for Specific Purposes (ESP) instruction in vocational education, where words refer to the objects and actions of a trade. Methodologically, it brings expert validity, classroom practicality, and effectiveness together in one 4D development study, whereas earlier Indonesian AR studies mostly reported feasibility or usability.

## **Research Method**

This study used research and development (R&D) with the 4D model (Define, Design, Develop, Disseminate; Thiagarajan, 1974) and judged product quality by validity, practicality, and effectiveness (Plomp & Nieveen, 2013). In Define, observation, interviews, and school records at SMK Negeri 9 Padang (2026/2027 academic year) identified the problem, the learners (Grade XI Culinary students), the task (understanding and using recipe procedure texts), and 40 target words: ingredients and utensils or equipment (20), cooking actions (10), and measurements and food qualities (10). In Design, a 40-item vocabulary test was written from a blueprint, and a storyboard and first prototype (Draft I) were produced for



Android-compatible devices, built in Unity with the Vuforia Engine providing the AR function; the content centres on one recipe, pizza margherita. In Develop, three experts appraised Draft I and the test, and the revised Draft II was tried out by three English teachers and 20 students, which led to the final product (Draft III). In Disseminate, effectiveness was tested with a pre-experimental one-group pretest–posttest design (Arikunto, 2009) over six meetings of  $2 \times 45$  minutes: the pretest was given in Meeting 1 and the posttest at the end of Meeting 6, and students worked mainly in pairs with the AR materials and took the tests individually. The final package (application, user guide, and word list) was then handed to the school's English teachers.

Validity was judged by three experts (one English-department lecturer and two English teachers of the school), and practicality by three English teachers and 20 Grade XI Culinary students who were not in the experimental class. The effectiveness test drew on a population of 104 students in three Grade XI Culinary classes. The experimental class was XI Culinary 4 (35 students), selected by purposive sampling, but the final dataset contains 30 students absent at the pretest or the posttest, and the purposive-sampling criteria.

Four instruments were used. The media validation sheet had 21 statements in four aspects (display, 6; ease of use, 5; technical quality, 5; suitability for learning, 5) rated on a four-point scale, and the test validation sheet had seven indicators rated on a five-point scale. The practicality questionnaire had parallel teacher and student versions of 20 statements in five aspects (understanding and content coverage, benefits, ease of access, clear procedures, and visual display) rated on a five-point scale. The vocabulary test had 40 items covering four indicators of vocabulary mastery: word meaning (multiple choice, items 1–15) and word classification (multiple choice, items 16–23), which are mainly receptive, and vocabulary recall (arranging scrambled words into a recipe sentence, items 24–37) and word usage (writing a sentence with a given cooking word, items 38–40), which are mainly productive (Nation, 2022). Items were scored 1 or 0 and converted to a 0–100 score. Item validity was tested with the item–total Pearson correlation (critical  $r = .361$  for  $N = 30$ ,  $\alpha = .05$ ; Sugiyono, 2013) and reliability with Cronbach's alpha (above .70 acceptable; Arikunto, 2009).

Expert ratings were analysed with Aiken's  $V$  (Aiken, 1985),  $V = \sum s / [n(c - 1)]$ , where  $s$  is the rating minus the lowest score,  $n$  the number of raters (3), and  $c$  the number of rating categories (4 for the media sheet, 5 for the test sheet);  $V \geq .80$  was taken as highly valid,  $.60 \leq V < .80$  as moderately valid, and  $V < .60$  as less valid (thresholds set for this study). Practicality was the total score as a percentage of the maximum score (81–100% very practical, 61–80% practical; Riduwan, 2022), and the internal consistency of the student questionnaire was checked with Cronbach's alpha. For effectiveness, pretest and posttest scores were described with means, standard deviations, and the number of students reaching the minimum passing score (KKM) of 75. Normality was tested with the Shapiro–Wilk test (Field, 2018); a paired-samples  $t$ -test was planned for normally distributed differences and the Wilcoxon signed-rank test otherwise, with the  $t$ -test reported for comparison. Effect sizes were  $r = Z/\sqrt{N}$  for the Wilcoxon test (Fritz et al., 2012) and  $d_z$  for the  $t$ -test (Lakens, 2013), interpreted with Cohen's (2013) benchmarks. The normalized gain,  $g = (S_{\text{post}} - S_{\text{pre}}) / (S_{\text{max}} - S_{\text{pre}})$ , was interpreted with Hake's (1998) categories (high,  $g \geq .70$ ; medium,  $.30 \leq g < .70$ ; low,  $g < .30$ ). The null hypothesis of no pretest–posttest difference was rejected at  $p < .05$ . Item analysis was run in SPSS and the other analyses in Python (SciPy).

## Results and Discussion

### The AR Vocabulary Builder

The Define and Design stages produced the AR Vocabulary Builder, a mobile application built around one recipe, pizza margherita (Figure 1). The main menu (Figure 1a) offers three vocabulary categories, equipment chef, tool, and ingredients, and an AR menu that opens the camera. In the AR view (Figure 1b), each object appears as a three-dimensional model over the real scene, with its name and a short definition below it, and on-screen buttons support navigation and manipulation of the model. Two assessment components complete the application: a 20-item multiple-choice exercise on procedure texts (Figure 1c) and a 10-item recipe-completion quiz in which students complete the steps of a pizza recipe with words from a word bank under a countdown timer (Figure 1d).



**Figure 1. Screenshots of the AR Vocabulary Builder**

*Note.* (a) Main menu with the three vocabulary categories. (b) AR view of a 3D object with its name and definition. (c) Multiple-choice exercise (item 1 of 20). (d) Recipe-completion quiz with a word bank (10 items).

### Validity

Table 2 shows the experts' ratings of the medium. The overall Aiken's  $V$  was .91 (93.3% of the maximum score), which falls in the highly valid category, and all four aspects reached  $V \geq .84$ . The display and the suitability of the medium for learning were rated highest ( $V = .96$  for both), whereas technical quality ( $V = .84$ ) and ease of use ( $V = .87$ ) were rated lowest. At item level,  $V$  ranged from .78 to 1.00; the four lowest values ( $V = .78$ ) concerned independent use by students, the response of the medium to user commands, the absence of disruptions, and the functioning of sound, video, animation, camera, and images.

**Table 2. Experts' Validation of the AR Vocabulary Builder and Vocabulary Test Using Aiken's  $V$**

| Aspect                   | Statements | M    | V   | % of maximum | Category     |
|--------------------------|------------|------|-----|--------------|--------------|
| Display                  | 6          | 3.89 | .96 | 97.2         | Highly valid |
| Ease of use              | 5          | 3.60 | .87 | 90.0         | Highly valid |
| Technical quality        | 5          | 3.53 | .84 | 88.3         | Highly valid |
| Suitability for learning | 5          | 3.87 | .96 | 96.7         | Highly valid |
| Medium, overall          | 21         | 3.73 | .91 | 93.3         | Highly valid |



|                                 |          |             |            |             |                     |
|---------------------------------|----------|-------------|------------|-------------|---------------------|
| <b>Vocabulary test, overall</b> | <b>7</b> | <b>4.71</b> | <b>.93</b> | <b>94.3</b> | <b>Highly valid</b> |
|---------------------------------|----------|-------------|------------|-------------|---------------------|

*Note.*  $N = 3$  experts. The medium was rated on a four-point scale and the vocabulary test (seven indicators) on a five-point scale.  $V =$  Aiken's (1985) index; categories follow the criteria given in the Method.

The three validators also judged the vocabulary test (Table 2). The overall  $V$  was .93 (94.3%), which is also highly valid. Consistency with the learning outcomes, the availability of a test blueprint, and the ease of understanding of the language received  $V = 1.00$ , whereas relevance to vocational learning in the culinary program and the clarity of the assessment indicators had the lowest value ( $V = .83$ ), still above the .80 criterion.

### Practicality

Table 3 presents the practicality ratings. The three teachers gave an overall mean of 4.55 (91.0%). The 20 students gave an overall mean of 4.45 (89.0%). Both groups therefore placed the medium in the very practical category, and every aspect was very practical for both groups. Teachers rated the visual display highest (100.0%, every rating being 5) and gave lower values to the benefits of the medium and to clear procedures (both 86.7%); their lowest-rated statements ( $M = 4.00$ ) concerned whether the medium helps students understand the learning objectives, improve their vocabulary mastery, and become motivated to expand their vocabulary. Students' ratings were even across aspects (87.5% to 89.8%). Their highest-rated statement was that the medium is easy to understand ( $M = 4.70$ ), and the lowest was that its features are easy to access ( $M = 4.25$ ). The internal consistency of the student questionnaire was high (Cronbach's  $\alpha = .90$ ).

**Table 3. Practicality of the AR Vocabulary Builder According to Teachers and Students**

| Aspect                                    | Teachers M  | Teachers %  | Students M  | Students %  | Category              |
|-------------------------------------------|-------------|-------------|-------------|-------------|-----------------------|
| Usage: understanding and content coverage | 4.67        | 93.3        | 4.48        | 89.5        | Very practical        |
| Benefits of the medium                    | 4.33        | 86.7        | 4.38        | 87.5        | Very practical        |
| Usage: ease of access                     | 4.42        | 88.3        | 4.48        | 89.5        | Very practical        |
| Clear procedures                          | 4.33        | 86.7        | 4.49        | 89.8        | Very practical        |
| Visual display                            | 5.00        | 100.0       | 4.43        | 88.5        | Very practical        |
| <b>Overall</b>                            | <b>4.55</b> | <b>91.0</b> | <b>4.45</b> | <b>89.0</b> | <b>Very practical</b> |

*Note.* Teachers  $n = 3$ ; students  $n = 20$ . Ratings used a five-point scale. Percentages are the total score as a percentage of the maximum possible score; categories follow the criteria given in the Method (Riduwan, 2022).

### Quality of the Vocabulary Test

Table 4 summarises the item analysis, which used the responses of the 30 students to the pretest. With  $N = 30$  (critical  $r = .361$ ), 33 of the 40 items were valid. The seven items that did not meet the criterion were item 14 ( $r = .182$ ), item 22 ( $r = .166$ ), item 23 ( $r = -.142$ ), item 24 ( $r = .356$ ), and the three sentence-writing items, 38 ( $r = .269$ ), 39 (correlation undefined because no student answered it correctly), and 40 ( $r = .208$ ). The sentence-writing items were answered correctly by very few students at the pretest (30.0%, 0.0%, and 3.3% for items 38, 39, and 40), a floor effect that restricts the correlation with the total score. Cronbach's alpha was .909 for the 40 items, which indicates high reliability, and .919 when the seven invalid items were excluded. All 40 items were retained in the pretest–posttest analysis for two reasons: the validators had judged the test highly valid in content ( $V = .93$ ), and removing the three sentence-writing items would have removed the word-usage indicator altogether. Because the item analysis drew on the pretest responses of the same students



rather than an independent try-out, the test was used unrevised in the posttest; items 14, 22, 23, and 24 should be reworded or replaced in the next version.

**Table 4. Item Analysis of the Vocabulary Test by Indicator**

| Indicator                            | Item numbers | Items     | Valid items | Invalid item no. | Pretest correct (%) |
|--------------------------------------|--------------|-----------|-------------|------------------|---------------------|
| Word meaning                         | 1–15         | 15        | 14          | 14               | 88.4                |
| Word classification                  | 16–23        | 8         | 6           | 22, 23           | 75.4                |
| Vocabulary recall (word arrangement) | 24–37        | 14        | 13          | 24               | 66.4                |
| Word usage (sentence writing)        | 38–40        | 3         | 0           | 38, 39, 40       | 11.1                |
| <b>Total</b>                         | <b>1–40</b>  | <b>40</b> | <b>33</b>   |                  | <b>72.3</b>         |

*Note.*  $N = 30$ . An item was valid when the Pearson correlation between the item and the total score exceeded the critical  $r$  of .361 ( $\alpha = .05$ ,  $df = 28$ ). Pretest correct (%) is the mean percentage of the items in the indicator answered correctly.

### Effectiveness

Table 5 shows that the mean score rose from 72.33 ( $SD = 18.30$ ) at the pretest to 87.93 ( $SD = 10.42$ ) at the posttest, a mean gain of 15.60 points. Twenty-nine of the 30 students improved, and the score of the remaining student changed by  $-0.5$  points (92.5 to 92). The number of students at or above the KKM of 75 rose from 20 (66.7%) to 28 (93.3%), and the number scoring 90 or higher rose from 3 to 18. The spread of scores also narrowed, as the standard deviation fell from 18.30 to 10.42.

**Table 5. Descriptive Statistics of the Vocabulary Test Scores**

| Score                            | M            | SD           | Mdn          | Min         | Max         |
|----------------------------------|--------------|--------------|--------------|-------------|-------------|
| Pretest                          | 72.33        | 18.30        | 78.75        | 20.0        | 92.5        |
| Posttest                         | 87.93        | 10.42        | 90.00        | 45.0        | 99.0        |
| <b>Gain (posttest – pretest)</b> | <b>15.60</b> | <b>10.95</b> | <b>10.00</b> | <b>-0.5</b> | <b>50.0</b> |

*Note.*  $N = 30$ . Scores range from 0 to 100.

The Shapiro–Wilk tests showed that the pretest, posttest, and difference scores all departed from normality (pretest  $W = .835$ , posttest  $W = .746$ , difference scores  $W = .828$ ; all  $p < .001$ ), so the Wilcoxon signed-rank test was the primary analysis. It showed a significant difference between the pretest and posttest scores,  $Z = 4.77$ ,  $p < .001$ , with a large effect ( $r = .87$ ). *The null hypothesis was therefore rejected.* The paired-samples  $t$ -test, reported for comparison, led to the same conclusion,  $t(29) = 7.81$ ,  $p < .001$ ,  $MD = 15.60$ , 95% CI [11.51, 19.69],  $dz = 1.43$  (Hedges'  $g_z = 1.39$ ). The normalized gain was .56 for the class, which is in the medium category, and the individual gains averaged .57 ( $SD = .19$ ): 6 students showed a high gain, 22 a medium gain, and 2 a low gain. Students who started lowest gained most ( $r = -.86$  between pretest score and gain); for example, the three lowest pretest scores (20.0, 25.0, and 47.5) rose to 70, 45, and 90, respectively.

### Discussion

#### Validity of the Product

The experts judged both the medium ( $V = .91$ ) and the vocabulary test ( $V = .93$ ) to be highly valid. The highest ratings for display and for suitability for learning suggest that the design choices of the Define and Design stages, which tied a small set of culinary words to three-dimensional objects and to the recipe context, were recognised as appropriate for the curriculum and the learners. This fits the situated view of vocabulary learning, in which words are learned together with the objects and activities they refer to (Brown et al., 1989; Santos et al., 2016), and the multimedia principle that words and pictures together support



learning when the design respects the limits of working memory (Mayer, 2020, 2024). The lowest values were concentrated in technical quality, notably the response of the medium to commands and the absence of disruptions, and in independent use by students. These items indicate where the next version should be improved before wider adoption. Agreement among only three experts is a modest basis for a validity claim, and the thresholds used to interpret V were set for this study.

### **Practicality of the Product**

Teachers (91.0%) and students (89.0%) both rated the medium very practical. Teachers rated the display highest but were more reserved about whether the medium helps students understand the learning objectives, improve their vocabulary, and stay motivated. This is understandable, because these statements concern outcomes that a short trial cannot fully show and that the effectiveness test addresses directly. Students' ratings were even across aspects, with ease of understanding and clear instructions rated highest, which matters in a school where kitchen practice is limited by having only one kitchen. The lowest-rated student statement concerned how easily the features can be accessed, a point to check in future versions. These results agree with earlier Indonesian development studies that found AR vocabulary media feasible and convenient for learners (Agata et al., 2021), and they add ratings from both teachers and students in a vocational setting. Because the ratings came from three teachers and 20 students who used the medium briefly, they describe perceived practicality rather than long-term classroom use.

### **Effectiveness of the Product**

Vocabulary scores increased significantly after six meetings with the AR Vocabulary Builder, and the effect was large by conventional benchmarks (Cohen, 2013). The result is consistent with studies reporting that AR improves EFL learners' vocabulary performance and motivation compared with conventional or flashcard-based teaching (Sadikin & Martyani, 2020; Tsai, 2020), and with reports of stronger engagement and gains among young learners and university ESP students (Amin, 2025; Hudaya & Sadikin, 2019; Lai & Chang, 2021). Three mechanisms proposed in the literature may help explain it: words and images presented together (Mayer, 2020, 2024), words learned in a context of objects and tasks (Santos et al., 2016), and active engagement in technology-based tasks (Kearsley & Shneiderman, 1998). This study did not test these mechanisms.

The evidence for AR is not uniform, and the present design cannot show that AR itself caused the gains. A meta-analysis of English vocabulary studies found an effect that favoured AR but was not statistically significant (Direkci & Aydemir, 2026), and immediate learning outcomes did not differ significantly between AR and non-AR users in Santos et al. (2016). Here there was no control group, the same 40-item test was used before and after, and the treatment combined AR objects with an exercise, a quiz, worksheets, and a presentation task, so the contribution of the AR component cannot be separated from that of the rest of the package. Testing and maturation effects remain possible. The gains were also largest for students who started lowest, which is expected when scores are bounded and may partly reflect regression toward the mean, and 18 of the 30 students scored 90 or more at the posttest, so a ceiling effect limits what the test could show for stronger students.

Several factors may help explain why the effect was so large ( $r = .87$ ), although the design cannot separate them. First,  $r = Z/\sqrt{N}$  reflects how consistently students improved (29 of 30) more than how much they improved: the mean gain was 15.60 points and the normalized gain only medium ( $g = .56$ ). Second, learner characteristics: students who started lowest gained most ( $r = -.86$  between pretest score and gain), so the size of the average gain



depends on where students start. Third, content: half of the 40 target words (20) named concrete objects, such as ingredients and utensils or equipment, that are easy to picture and to link to a three-dimensional model, and they were taught and tested with closely aligned materials over six meetings. Fourth, treatment: the package combined AR with an exercise, a quiz, worksheets, a presentation task, pair work, and teacher guidance, and the novelty of AR may have raised attention in a short intervention. Fifth, context: with only one kitchen, students may have had little prior hands-on exposure to the target objects, so the AR view may add more here than where real objects are at hand. Which of these moderated the effect is an empirical question for a controlled study that varies learners' baseline level, treatment components, and duration.

At the pretest, students did best on word meaning (88.4% correct) and worst on word usage (11.1%), with word classification (75.4%) and word arrangement (66.4%) in between. This gradient is consistent with the distinction between receptive and productive vocabulary knowledge (Nation, 2022), and it suggests that the most room for improvement lay in productive use. Because item-level posttest data were not available for this analysis, it is not known which indicator changed most. The main limitations are the one-group design without a control group, the small sample from a single school, an immediate posttest without a retention check, and a medium that covers one recipe.

For vocational teachers with limited access to practical kitchen facilities, the AR Vocabulary Builder can serve as a virtual kitchen for the language side of recipe work. At SMK Negeri 9 Padang, which has only one kitchen, teachers can use the application before kitchen sessions, so that students already know the names of equipment, ingredients, and actions when they enter the kitchen and scarce kitchen time goes to practice rather than vocabulary explanation; run it in class on students' own Android-compatible devices, in pairs where devices are scarce; and use the exercise and the recipe-completion quiz as quick checks that need no special equipment. Because the lowest-rated student statement concerned how easily the features can be accessed, schools should check device availability and the technical response of the application before use. The study did not test whether better vocabulary transfers to performance in the kitchen, so the AR medium should complement, not replace, hands-on practice.

## **Conclusion**

This study developed an AR-based Vocabulary Builder on recipe procedure text for Grade XI Culinary students with the 4D model. The product was highly valid according to three experts ( $V = .91$ ), and it was rated very practical by three teachers (91.0%) and 20 students (89.0%). In the effectiveness test with 30 students, vocabulary scores rose significantly from a mean of 72.33 to 87.93 (Wilcoxon  $Z = 4.77$ ,  $p < .001$ ,  $r = .87$ ), with a medium normalized gain ( $g = .56$ ). The findings support the AR Vocabulary Builder as a valid, practical, and promising medium for vocabulary learning in vocational EFL classes. Because of the one-group design, the ceiling effect at the posttest, and the small sample, the effect on vocabulary mastery should be confirmed in a controlled study.

## **Recommendation**

For teachers: use the application as a supplement to recipe-procedure lessons, in pairs when devices are limited, and add productive tasks such as sentence writing and recipe completion, because word usage was the weakest indicator at the pretest (11.1% correct). For school principals and curriculum developers: provide devices and technical support for AR



use, since technical quality received the lowest expert ratings ( $V = .84$ ) and ease of access the lowest-rated student statement, and consider extending the content to further recipes and other competencies of the culinary program. For future researchers: use a quasi-experimental design with a control group and a delayed posttest; analyse results by indicator; revise or replace the invalid and sentence-writing items and try the test out independently of the experimental class; separate the contribution of the AR component from that of the rest of the package; and test the possible moderators discussed above in larger, multi-school samples.

## References

- Agata, D., Yuniarti, H., & Adison, A. A. P. (2021). Teaching English Vocabulary to Young Learners via Augmented Reality Learning Media. *Beyond Words*, 9(2), 91–99. <https://doi.org/10.33508/bw.v9i2.2772>
- Aiken, L. R. (1985). Three Coefficients for Analyzing the Reliability and Validity of Ratings. *Educational and Psychological Measurement*, 45(1), 131–142. <https://doi.org/10.1177/0013164485451012>
- Amin, E. A.-R. (2025). The Impact of Augmented Reality on Vocabulary Acquisition and Engagement Among College ESP Students. *Theory and Practice in Language Studies*, 15(2), 414–425. <https://doi.org/10.17507/tpls.1502.11>
- Arikunto, S. (2009). *Dasar-dasar evaluasi pendidikan (Edisi Revisi)*. Bumi Aksara. Perpustakaan Sekolah Tinggi Ilmu Kepolisian (STIK). <https://library.stik-ptik.ac.id>
- Azuma, R., Baillot, Y., Behringer, R., Feiner, S., Julier, S., & MacIntyre, B. (2001). Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6), 34–47. <https://doi.org/10.1109/38.963459>
- Belda-Medina, J., & Marrahi-Gomez, V. (2023). The Impact of Augmented Reality (AR) on Vocabulary Acquisition and Student Motivation. *Electronics*, 12(3), 749. <https://doi.org/10.3390/electronics12030749>
- Binhomran, K., & Altalhab, S. (2024). The impact of implementing augmented reality to enhance the vocabulary of young EFL learners. *The JALT CALL Journal*, 17(1), 23–44. <https://doi.org/10.29140/jaltcall.v17n1.304>
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated Cognition and the Culture of Learning. *Educational Researcher*, 18(1), 32–42. <https://doi.org/10.3102/0013189X018001032>
- Buchner, J., Buntins, K., & Kerres, M. (2022). The impact of augmented reality on cognitive load and performance: A systematic review. *Journal of Computer Assisted Learning*, 38(1), 285–303. <https://doi.org/10.1111/jcal.12617>
- Cohen, J. (2013). *Statistical Power Analysis for the Behavioral Sciences*. Routledge.
- Direkci, B., & Aydemir, N. (2026). Augmented Reality Technology in Language Learning: A Meta-Analysis of Experimental Studies (2015–2025). *International Journal of Learning, Teaching and Educational Research*, 25(3). <https://doi.org/10.26803/ijlter.25.3.21>
- Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (Vol. 5). sage London.
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Gerot, L., & Wignell, P. (2015). *Making Sense of Functional Grammar*. Gerd Stabler. <https://openlibrary.telkomuniversity.ac.id/home/catalog/id/104961/slug/making-sense-of-functional-grammar.html>



- Hadi, S. H. (2025). Development of Augmented Reality-Based English Learning Media for Vocabulary and Pronunciation. *JEECS (Journal of Electrical Engineering and Computer Sciences)*, 10(2), 152–159. <https://doi.org/10.54732/jeeecs.v10i2.6>
- Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. *American Journal of Physics*, 66(1), 64–74. <https://doi.org/10.1119/1.18809>
- Hartini, L. W., & Ardini, A. S. (2024). The Relationship Between Indonesian EFL Learners' Vocabulary Knowledge and English Competencies. *Journal of English and Education (JEE)*, 55–66. <https://doi.org/10.20885/jee.v10i1.33651>
- Hudaya, A., & Sadikin, I. S. (2019). The effect of aurasma augmented reality (AR) to enhance young learners' vocabulary mastery. *PROJECT (Professional Journal of English Education)*, 2(6), 783–790.
- Kearsley, G., & Shneiderman, B. (1998). Engagement Theory: A Framework for Technology-Based Teaching and Learning. *Educational Technology*, 38(5), 20–23. <https://www.jstor.org/stable/44428478>
- Khan, R. M. I., Ali, A., Kumar, T., & Venugopal, A. (2023). Assessing the efficacy of augmented reality in enhancing EFL vocabulary. *Cogent Arts & Humanities*, 10(1), 2223010. <https://doi.org/10.1080/23311983.2023.2223010>
- Korosidou, E. (2024). The Effects of Augmented Reality on Very Young Learners' Motivation and Learning of the Alphabet and Vocabulary. *Digital*, 4(1), 195–214. <https://doi.org/10.3390/digital4010010>
- Kosniawati, Fitriani, D., Aziz, U. B. A., Azizah, & Andriansyah. (2026). Developing augmented reality flashcards for English vocabulary instruction in Indonesian early childhood education: An ADDIE model feasibility study. *Englisia: Journal of Language, Education, and Humanities*, 13(2), 658–676. <https://doi.org/10.22373/englisia.173>
- Krüger, J. M., & Bodemer, D. (2022). Application and Investigation of Multimedia Design Principles in Augmented Reality Learning Environments. *Information*, 13(2), 74. <https://doi.org/10.3390/info13020074>
- Lai, J.-Y., & Chang, L.-T. (2021). Impacts of Augmented Reality Apps on First Graders' Motivation and Performance in English Vocabulary Learning. *Sage Open*, 11(4), 21582440211047549. <https://doi.org/10.1177/21582440211047549>
- Lakens, D. (2013). Calculating and reporting effect sizes to facilitate cumulative science: A practical primer for t-tests and ANOVAs. *Frontiers in Psychology*, 4. <https://doi.org/10.3389/fpsyg.2013.00863>
- Larchen Costuchen, A., Darling, S., & Uytman, C. (2021). Augmented reality and visuospatial bootstrapping for second-language vocabulary recall. *Innovation in Language Learning and Teaching*, 15(4), 352–363. <https://doi.org/10.1080/17501229.2020.1806848>
- Marrahí-Gómez, V., & Belda-Medina, J. (2023). The Effect of Using AR Technology on Language Learning. *International Journal of English Language Studies*, 5(1), 53–58. <https://doi.org/10.32996/ijels.2023.5.1.5>
- Maula, M. (2021). The development of English learning materials for culinary art program using four-d model. *RETAIN: Journal of Research in English Language Teaching*, 9(2), 10–18.
- Mayer, R. E. (2020, July 9). *Multimedia Learning*. Cambridge University Press. Cambridge Aspire Website. <https://doi.org/10.1017/9781316941355>



- Mayer, R. E. (2024). The Past, Present, and Future of the Cognitive Theory of Multimedia Learning. *Educational Psychology Review*, 36(1), 8. <https://doi.org/10.1007/s10648-023-09842-1>
- Nabujing, N. H., & Utari, P. (2023). Writing Procedure Text by Using Team Pair Solo Technique. *EXCELLENCE: Journal of English and English Education*, 3(2), 52–57. <https://doi.org/10.47662/ejeee.v3i2.652>
- Nation, I. (2006). How Large a Vocabulary is Needed For Reading and Listening? *The Canadian Modern Language Review*, 63(1), 59–82. <https://doi.org/10.3138/cmlr.63.1.59>
- Nation, I. S. P. (2022). *Learning Vocabulary in Another Language* (3rd edn). Cambridge University Press. <https://doi.org/10.1017/9781009093873>
- Plomp, T., & Nieveen, N. (2013). *Educational design research*.
- Radu, I., & Schneider, B. (2023). How Augmented Reality (AR) Can Help and Hinder Collaborative Learning: A Study of AR in Electromagnetism Education. *IEEE Transactions on Visualization and Computer Graphics*, 29(9), 3734–3745. <https://doi.org/10.1109/TVCG.2022.3169980>
- Riduwan, M. (2022). *Skala pengukuran variabel-variabel penelitian*.
- Sadikin, I. S., & Martyani, E. (2020). Integrating Augmented Reality (AR) In EFL Class For Teaching Vocabulary | PROJECT (Professional Journal of English Education). <https://journal.ikipsiliwangi.ac.id/index.php/project/article/view/3841>
- Santos, M. E. C., Lübke, A. in W., Taketomi, T., Yamamoto, G., Rodrigo, Ma. M. T., Sandor, C., & Kato, H. (2016). Augmented reality as multimedia: The case for situated vocabulary learning. *Research and Practice in Technology Enhanced Learning*, 11(1), 4. <https://doi.org/10.1186/s41039-016-0028-2>
- Sugiyono. (2013). *Metode Penelitian Pendidikan Pendekatan Kuantitatif, Kualitatif dan R&D*. Alfabeta. (Bandung).  
[//digilib.unigres.ac.id%2Findex.php%3Fp%3Dshow\\_detail%26id%3D43](http://digilib.unigres.ac.id%2Findex.php%3Fp%3Dshow_detail%26id%3D43)
- Thiagarajan, S. (1974). *Instructional development for training teachers of exceptional children: A sourcebook*. <http://files.eric.ed.gov/fulltext/ED090725.pdf>
- Tsai, C.-C. (2020). The Effects of Augmented Reality to Motivation and Performance in EFL Vocabulary Learning. *International Journal of Instruction*, 13(4), 987–1000. <https://doi.org/10.29333/iji.2020.13460a>
- Wilkins, D. A. (1972). *Linguistics in language teaching* (Vol. 111). Edward Arnold London.