

The Effectiveness of the V.A.R.K Model in Acquiring Grammatical Concepts Among Eighth-Grade Students in Jordan

Hamza Moteb Al-Zuhair¹, Raed Mahmoud Khodair²

¹Department of Curriculum and Instruction, Faculty of Educational Sciences, Yarmouk University, Jordan
Email: alzhryhmzt@gmail.com

ORCID: <https://orcid.org/0009-0006-1310-2833>

²Department of Curriculum and Instruction, Faculty of Educational Sciences, Yarmouk University, Jordan
Email: raedk@yu.edu.jo

ORCID: <https://orcid.org/0009-0001-0281-9998>

*Corresponding author: Raed Mahmoud Khodair, E-mail: raedk@yu.edu.jo

HOW TO CITE:

Hamza Moteb Al-Zuhair, Raed Mahmoud Khodair (2026). The Effectiveness of the V.A.R.K Model in Acquiring Grammatical Concepts among Eighth-Grade Students in Jordan. International Journal of Special Education, 41(11s), 1006-1018

COPYRIGHT STATEMENT:

Copyright: © 2026 Authors.
Open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

ABSTRACT:

Arabic grammar instruction in Jordanian middle schools emphasizes rule memorization and decontextualized learning, yet assessments show low proficiency among eighth-grade students. This study aims to examine whether an eight-week V.A.R.K.-based instructional program improves immediate receptive grammatical knowledge in eighth-grade male students compared to traditional instruction. Fifty-four male students (experimental $n = 27$, control $n = 27$) participated in a quasi-experiment. The experimental group received 12 hours of instruction incorporating visual, auditory, read/write, and kinesthetic activities based on the V.A.R.K. model, Cognitive Load Theory, and Skill Acquisition Theory. Both groups completed pre- and post-tests; ANCOVA used pre-test scores as covariates, with bootstrap (1,000 iterations) for 95% CI and sensitivity analysis. ANCOVA showed a significant group effect, $F(1, 51) = 38.32$, $p < .001$, $\eta^2 = .429$, 95% CI [.312, .511]; Hedges' $g = 1.84$, 95% CI [1.14, 2.54]. Sensitivity analyses after removing outliers confirmed stability ($\eta^2 = .426$). Productive grammar scores were lower than receptive scores with no differential trend. An eight-week V.A.R.K.-based instructional program supports short-term receptive acquisition of Arabic grammatical concepts.

Keywords: V.A.R.K. model, Arabic grammar, receptive grammatical knowledge, quasi-experimental study, Cognitive Load Theory.

1. Introduction

When nearly 60 percent of Jordan's eighth grade students are unable to recognize a predicate in a sentence — even though they have learned this basic element of grammar for just two years — questions need to be asked regarding how instruction has been designed, rather than what abilities students possess. The national assessment results from the Ministry of Education (2022) show that only 43 percent of

students can demonstrate grammatically correct use of Arabic, while specific areas such as predicates (31 percent) and the five nouns (28 percent) represent major weakness. Despite the implementation of the new curriculum reform emphasizing learner centered teaching methods and conceptual understanding of content (Al-Shammari, 2020), these persistent gaps exist. Arabic grammar (Nahw) represents one of the fundamental pillars of language learning (Al-Hawamdeh & Ashour, 2015). In the Jordanian

eighth-grade curriculum), students are required to develop understanding of increasingly complex structures, including the subject (Mubtada'), predicate (Khabar), the five nouns, and indeclinable verbs. Traditional instruction in Arabic grammar continues to emphasize rule-based teaching and mechanical practice, leading to rote memorization rather than conceptual understanding of grammatical structures (Al Zaatreh & Al Khawaldeh, 2024).

1.1 Theoretical Framework

This study is grounded in three complementary frameworks: the V.A.R.K. model, Cognitive Load Theory (Sweller, 1988), and Skill Acquisition Theory (DeKeyser, 2003). The V.A.R.K. model emphasizes the use of visual, auditory, read/write, and kinesthetic instructional activities to facilitate learning. Cognitive Load Theory explains how structured instructional design can reduce unnecessary cognitive processing during complex tasks, while Skill Acquisition Theory describes the progression from declarative knowledge to procedural fluency through guided practice. Together, these frameworks suggest that V.A.R.K.-based instructional environments may support grammatical concept acquisition more effectively than rule-based instruction. Selected constructivist principles—collaborative sentence construction, guided grammatical categorization, and kinesthetic sorting—operationalize active participation in grammatical meaning-making processes.

1.2 The Empirical Gap

Although numerous studies have demonstrated that multi-sensory methods can improve students' performance with regards to learning a second language (Abdul Kareem & Hassan, 2023; Al-Zaatreh & Al-Khawaldeh, 2024; Nurhadi & Hilmi, 2022), this is true of all studies based on the Middle East region. Unfortunately, there has been little work done within the specific context of the Arab world that examines whether or not

these methods can be used as an effective means for improving the academic success of elementary school-aged students. To our knowledge, no prior study has combined V.A.R.K.-based Arabic grammar instruction with a quasi-experimental pilot design incorporating fidelity monitoring, differentiated receptive and exploratory productive assessment, and robustness-oriented statistical procedures.

1.3 The Present Study

Accordingly, this pilot investigation addresses the following research questions: *RQ1*: Does V.A.R.K.-based instruction produce a statistically significant effect on eighth-grade students' immediate receptive grammatical recognition compared to traditional instruction?

RQ2: Does the effect of V.A.R.K.-based instruction differ across grammatical subdomains (subject, predicate, the five nouns, indeclinable verbs)?

RQ3 (exploratory): Do productive grammatical tasks demonstrate different performance patterns relative to receptive recognition following the intervention?

Based on the theoretical frameworks, we hypothesized that students receiving V.A.R.K.-based instruction would demonstrate significantly higher immediate post-test receptive grammatical recognition than students receiving traditional instruction after controlling for pre-test performance (H1), and that instructional effects would differ across grammatical subdomains (H2).

The study is explicitly framed as a pilot quasi-experimental investigation designed to generate preliminary effect estimates, evaluate instructional feasibility, and identify methodological considerations relevant to future large-scale randomized studies. The results pertain only to short-term recognition outcomes in a single-school, all-male group and should be interpreted as exploratory and hypothesis-generating rather than as conclusive evidence of causal superiority or population-level generalizability.

2. Literature Review

A systematic literature review was conducted to identify empirical and theoretical studies related to Arabic grammar instruction, the V.A.R.K. model, cognitive load, and grammar acquisition. Academic databases including Google Scholar, Scopus, and ERIC were searched, in addition to Arabic-language educational databases. The search focused on studies published between 2015 and 2024 to capture recent developments in Arabic language instruction. Search terms included “Arabic grammar instruction,” “grammar acquisition,” “multimodal learning,” “V.A.R.K.-based instruction,” “learning styles critique,” “cognitive load,” “grammatical concepts,” and “V.A.R.K. model.” Studies were selected if they examined the effectiveness of multimodal approaches in grammar teaching or language acquisition (Ellis, 2006). Only experimental and quasi-experimental studies conducted in elementary or middle-school educational contexts were included.

2.1 Arabic Grammar Instruction: Persistent Challenges

The study of Arabic syntax (Arabic Nahw) involves learning how words are related to one another and applying the patterns of language to create meaningful content (Hasan, 2011). Most of today’s instruction at schools in Jordan is still very much a traditional teacher-led method with an emphasis on rote memorization rather than conceptual understanding (Al-Shammari, 2020; Wahba, Taha, & Giolfo, 2023). At the middle school level, students will be required to learn the compound structures for example, the subject (mubtada') and predicate (khabar), the five noun types and the indeclinable verb—all of which require critical thinking skills such as analysis, classification and application of structure within context (Khasawneh, 2022).

Arabic is a language with special cognitive demands in its use of structure as part of the basis for evaluating the effectiveness of teaching students about structured grammar. Because Arabic's (as opposed to most

other languages) syntax does not depend largely on fixed word order, learners are required to pay attention to syntactic placement, morphological marking, case endings, relationships of subject-verb-object agreement and contextual meanings. As such, the flexibility of the word order in Arabic and the increased importance of both morphological and syntactical cues increase the amount of cognitive load that is associated with the acquisition of Arabic by a typical middle school student. Therefore, instructional methods which are limited solely to explaining rules or memorizing them will likely fall short in providing an opportunity for students to develop their own conceptual understanding of how grammar works. V.A.R.K.-based strategies may support learning by making abstract syntactic relations more visible, discussable, and manipulable through sentence mapping, classification tasks, oral explanation, and contextualized practice.

2.2 Theoretical Frameworks for V.A.R.K.-Based Instruction

Contemporary instructional theory emphasizes V.A.R.K.-based instructional approaches supported by complementary cognitive frameworks (Mekki, 2023). The V.A.R.K. model incorporates visual, auditory, read/write, and kinesthetic learning activities to facilitate learning. Cognitive Load Theory (Sweller, 1988) explains how instructional design may reduce unnecessary processing demands during complex learning tasks, while Skill Acquisition Theory (DeKeyser, 2003) conceptualizes grammatical learning as a gradual transition from declarative rule awareness toward proceduralized performance through guided practice. Research in multimedia learning similarly suggests that combining visual and verbal instructional representations may improve conceptual processing when materials are cognitively well-structured (Mayer, 2021).

2.2.1 Critique of Learning Styles Model

The present study adopts the V.A.R.K. model as an instructional framework incorporating visual, auditory, read/write, and kinesthetic learning activities. Although the V.A.R.K. model has been criticized within the learning styles literature, Willingham et al. (2015) demonstrated that no rigorous evidence supports matching instructional modalities to claimed sensory learning styles, and Nancekivell et al. (2020) found that learning styles beliefs function as essentialist "folk psychology" rather than scientifically validated constructs. Accordingly, the present study does not interpret the V.A.R.K. model as a mechanism for matching instruction to individual sensory preferences. Rather, it is used as a structured instructional framework that incorporates multiple learning activities to enhance engagement and conceptual understanding during grammar instruction.

2.2.2 Integrated Framework

Taken together, the V.A.R.K. model, Cognitive Load Theory, and Skill Acquisition Theory provide an integrated explanatory basis for the present intervention. The V.A.R.K. model provides a structured framework that incorporates visual, auditory, read/write, and kinesthetic learning activities.

Cognitive Load Theory examines how a learner's ability to utilize multiple (visual, auditory, written, kinesthetic) modalities in order to structure learning aids can decrease the amount of additional, or unnecessary processing required by a learner when encountering new abstract grammar concepts. Skill Acquisition Theory also explains that repeated exposure to the same guided practices can lead to a shift away from declarative awareness of rules toward more automatic fluency with regard to recognizing grammatical patterns. Therefore, the V.A.R.K. model provides the instructional framework, Cognitive Load Theory addresses cognitive efficiency in the learning process, and Skill Acquisition Theory describes the developmental progression through which learners access and utilize previously acquired grammatical knowledge.

2.3 Empirical Evidence on V.A.R.K.-Based Grammar Instruction

Most studies report positive results from using V.A.R.K.-based and multimodal teaching approaches, as opposed to a single instructional modality, for language learning (Abdul Kareem & Hassan, 2023; Nurhadi & Hilmi, 2022; Brosh, 2017; Al-Zaatreh & Al-Khawaldeh, 2024).

Table 1. Summary of Selected Studies on V.A.R.K.-Based and Multimodal Grammar Instruction

Study	Design	Sample	Theoretic al Framewor k	Fidelity Monitorin g	Delaye d Post- Test	Productiv e Assessme nt	Major Limitatio n
Abdul Kareem & Hassan (2023)	Quasi- experiment al	62 primary students	V.A.R.K. model	No	No	No	Learning styles framewor k
(Nurhadi & Hilmi, 2022)	Pretest- posttest	48 secondar y students	V.A.R.K. model	No	No	No	Small sample; learning styles framewor k

(Brosh, 2017)	Experimental	School students	Multimodal	No	No	Limited	Productive skills not assessed
Al-Zaatreh & Al-Khawaldeh (2024)	Quasi-experimental	Jordanian eighth-grade	Instructional scaffolding	No	No	No	No multimodal integration
Present study	Pilot quasi-experimental	54 eighth-grade	V.A.R.K. model + CLT + SAT	Yes (92% IRR)	No	Exploratory ($\alpha=.81$)	Single-school pilot sample

The studies cited above have found similar results with regard to concept mapping, gamified grammar teaching, and multimodal scaffolding in Arabic language teaching and learning across Egypt, Saudi Arabia, and other countries in the region. Gamification was found to increase student motivation, interest, and academic success in Arabic vocabulary and grammar learning (Almelhes, 2024). On an international level, Yu et al. (2026), presented a review of empirical research on multimodal composing using digital media in L2 classroom contexts and reported overall positive results concerning learner engagement, meaning making, and language performance.

Although the empirical research has shown mostly positive results related to multilingual learners and multimodal teaching methods, the current body of empirical research has consistently demonstrated three major methodological limitations. First, most of the studies conducted to date do not include a delayed posttest design. Therefore, the majority of studies are unable to determine if the use of multimodal instructional input will result in better long-term retention and application compared to a single modality. However, when researchers have included a delayed posttest into their study, they have found that multimodal instructional input may be more effective at promoting retention and application of content compared to

single-modality approaches (Zhang & Zhang, 2024).

As stated previously, this finding remains highly dependent upon the specific context in which the research is being conducted. Second, while most studies examine students' ability to recognize grammatical structures, few studies assess productive grammatical competence. This represents a significant limitation because most educators believe that students should be able to produce grammatically correct sentences as well as recognize them. Finally, although numerous critiques exist regarding the effectiveness of learning style-matching frameworks for interpreting how students learn from multimodal instruction, many studies continue to interpret multimodal instruction within learning style-based frameworks.

2.4 Gaps in the Literature and the Present Contribution

A review of the literature reveals multiple persistent gaps that need further exploration. Initially, there is a deficiency of quasi-experimental research carried out in Jordanian middle school settings that concentrate on Arabic grammar. Second, most studies lack fidelity monitoring, preventing assessment of implementation quality. Third, productive grammatical competence is rarely assessed. Fourth, few studies have integrated the

V.A.R.K. model alongside Cognitive Load Theory and Skill Acquisition Theory within Arabic grammar instruction.

To our knowledge, no prior study has combined V.A.R.K.-based Arabic grammar instruction with a quasi-experimental design incorporating fidelity monitoring, differentiated receptive and productive assessment, and robustness-oriented statistical procedures. The present study addresses these gaps by implementing a pilot quasi-experimental design grounded in the V.A.R.K. model, Cognitive Load Theory, and Skill Acquisition Theory, with fidelity monitoring (92% inter-rater reliability), exploratory productive assessment ($\alpha = .81$), and robustness checks including bootstrapping and sensitivity analysis.

3. Methodology

3.1 Research Design

A quasi-experimental pretest-posttest control-group design was used to evaluate the effectiveness of a V.A.R.K.-based instructional program for eighth-grade grammar acquisition in Jordan. Intact classrooms were randomly assigned to experimental or control conditions after confirming no significant differences in prior-year Arabic test scores ($t(52) = 0.44$, $p = .66$) or pre-test grammar scores ($t(52) = 0.70$, $p = .49$). This pilot investigation aimed to generate initial evidence and assess the feasibility of the instructional approach. The quasi-experiment utilized a quasi-experimental pre-test/post-test control group structure, where the treatment group initially participated in a pre-test (O1) and then underwent the multi-sensory instructional program (X), followed by a post-test (O2). The comparison/control group also underwent both the pre-test (O1) and post-test (O2), but did not receive the multi-sensory instructional program.

3.2 Participants and Sampling

A group of 54 male students in eighth grade was gathered from a public school situated in the Amman Governorate of Jordan. Participants were gathered via convenience sampling from existing classroom groups at the school involved. The researcher's institutional access to the school

provided a controlled environment for data collection, the ability to monitor intervention fidelity, and access to instructional scheduling, all of which were important for conducting the present pilot investigation. The school was an all-boys' school, consistent with many public secondary schools throughout Jordan. Generalization to girls would require additional research studies. The estimated power analysis (G*Power 3.1), based on the following parameters: $\alpha = .05$, Power = .80 and a moderate effect size $f = .25$, indicated a minimum of 48 participants should be included. Although the final sample included 54 participants, it is acknowledged that the current sample can only support the detection of large effects ($f \geq .40$). Additionally, the results of the pilot study are not representative of all Jordanian students in their last year of elementary school (Ministry of Education, Jordan, 2019).

3.3 Research Instrument

3.3.1 Grammatical Concepts Test

A research-based test was developed to measure students' learning of grammatical concepts as identified by the Jordanian Eighth-Grade Curriculum and other research that have previously studied these areas. The test consisted of thirty (30) multiple choice questions that covered the four main areas of grammar. These areas were subject (mubtada') and all of its forms (six items); predicate (khabar) and all of its forms (six items); the five nouns (six items); and indeclinable verbs (twelve items). Since indeclinable verbs are cognitively much more complex than the other three areas, it is a major part of Unit 3 of the curriculum (12 of the 30 instructional hours) (Ministry of Education, Jordan, 2020). Additionally, 3 constructed-response items and 1 error-analysis task formed an exploratory productive subscale ($\alpha = .81$).

3.3.2 Construct Validity

Exploratory Factor Analysis (Principal Axis Factoring with Promax Rotation) indicated one primary factor explaining 58% of the variance, which is acceptable for unidimensional constructs (Hair et al., 2019). Total scores were

used as the main measure ($\alpha = .96$), and the four constructed-response items formed a secondary productive subscale ($\alpha = .81$), rated by two independent evaluators (ICC = .87, 95% CI [.79, .92]). Instrument development and validation followed established psychometric principles (DeVellis & Thorpe, 2021).

3.4 Validity and Reliability Procedures

The content validity of the survey was determined by having the survey reviewed by ten experts (i.e., university faculty members, curriculum specialists). Based upon their recommendations, revisions were made to develop an item-scale Level Content Validity Index (Scale- S-CVI) that was .89. The individual-item level of CVI (Item-I-CVI) for all items developed during this process had values ranging from .70 to 1.00. Three items which did not have an I-CVI value greater than or equal to .78 were revised before they were administered as part of the study. Additionally, all items demonstrated acceptable construct validity because all item-total correlation values exceeded .30. All items also demonstrated adequate discriminative ability, since the p-value (difficulty index) values for each item ranged from .42 to .78 and all point-biserial discrimination coefficient values exceeded .35. Test-retest reliability was not evaluated due to the short time period of the intervention and possible practice effects.

3.5 Intervention Procedures

The intervention lasted eight weeks, with two 45-minute sessions per week (12 hours total). Both conditions received equal instructional time from the same teacher and followed the same instructional objectives. As the researcher delivered the intervention, potential instructional allegiance and expectancy effects could not be fully eliminated.

Experimental Group

Students received V.A.R.K.-based instruction incorporating visual, auditory, read/write, kinesthetic activities, and collaborative sentence construction.

Control Group

Students received traditional rule-based instruction. Fidelity of implementation was monitored to ensure the experimental techniques were not applied in the control group. Detailed weekly instructional matrices and sample lesson plans are available in supplementary materials

3.6 Fidelity of Implementation

Fidelity was monitored using a 10-item checklist (0–5). The educator received training in the V.A.R.K. instructional approach and observed practice sessions. A neutral observer evaluated 30% of sessions. Inter-rater reliability was 92% (Cohen's $\kappa = .89$). While the observer was aware of group conditions, the checklist minimized subjective bias.

3.7 Statistical Analysis

The analysis of data was performed utilizing SPSS 29.0. Assumption testing confirmed normality (Shapiro-Wilk $W > .95$ across groups), homogeneity of variance (Levene's $F = 1.92$, $p = .17$), homogeneity of regression slopes ($F(1,50) = 1.04$, $p = .31$), and equality of covariance matrices for MANCOVA (Box's $M = 12.45$, $p = .12$). Because failure to reject the null hypothesis cannot be used to verify assumptions, we also investigated Q-Q plots and residuals. The Q-Q plot and residuals indicated that there were no serious violations in either.

The primary outcome was analyzed using an Analysis of Covariance (ANCOVA), where the dependent variable is the post-test score, the independent variable is instructional condition, and the covariate is the pre-test score. Eta Squared (η^2) values for each effect size will be reported along with their corresponding 95% confidence interval (bootstrap with 1,000 samples). In addition to reporting η^2 values, Hedges' g (small sample correction) will be calculated for the primary outcome. A sensitivity analysis for outliers removed the highest 5% and lowest 5% of post-test scores. Since the purpose of domain level MANCOVA was exploratory in nature, Bonferroni correction was not applied to minimize Type II error; therefore, all results should be viewed from a descriptive standpoint.

4. Results

4.1 Descriptive Statistics and Baseline Equivalence

Table 2 provides descriptive statistics for grammatical performance before and after the

test across both instructional conditions. The experimental group ($n = 27$) and the control group ($n = 27$) showed similar baseline performance. (see Table 2).

Table 2. Descriptive Statistics and Baseline Equivalence

Group	n	Pre-test M (SD)	Post-test M (SD)	Adjusted Post-test M (SE)
Experimental	27	11.2 (3.1)	18.7 (2.8)	18.5 (0.45)
Control	27	10.8 (3.4)	12.4 (3.1)	12.6 (0.45)

Independent-samples t -tests indicated no significant baseline differences between the experimental and control groups in prior-year Arabic achievement, $t(52) = 0.44$, $p = .66$, or pre-test grammatical concept scores, $t(52) = 0.70$, $p = .49$. Furthermore, the correlation between pre-test and post-test scores was moderate to strong in both the experimental group ($r = .58$) and the control group ($r = .62$), supporting the use of ANCOVA and confirming that pre-test scores were an appropriate covariate for the analysis.

4.2 Primary Outcome: ANCOVA Results

A one-way ANCOVA was conducted to examine the effect of instructional condition on post-test grammatical performance after adjusting for pre-test scores. A significant difference was found between groups, with the experimental group outperforming the control group, $F(1,51) = 38.316$, $p < .001$, $\eta^2 = .429$, 95% CI [.312, .511], indicating a large effect (Cohen, 2013). The adjusted R^2 was .415, showing that the model explained 41.5% of the variance in post-test performance.

4.2.1 Hedges' g (correction for small samples)

To provide a standardized effect size appropriate for pilot studies, Hedges' g was calculated. The

analysis yielded a very large effect size ($g = 1.84$, 95% CI [1.14, 2.54]), indicating that the average post-test score of students in the experimental group was approximately 1.84 standard deviations higher than that of students in the control group after controlling for pre-test variation.

4.3 Robustness Analyses

4.3.1 Confidence intervals derived from bootstrapping

To assess the stability of the observed effect size, bootstrapped 95% confidence intervals for η^2 were computed using 1,000 resamples. The resulting confidence interval ranged from .312 to .511 and did not include zero, indicating that the effect was stable and unlikely to be attributable to sample-specific variation.

4.3.2 Sensitivity evaluation (influence of outliers)

To assess if extreme scores affected the results, we performed a sensitivity analysis by excluding the highest and lowest 5% of post-test scores ($n = 5$ total observations, roughly 9% of the sample). The re-analysis confirmed the pattern of results: $F(1, 46) = 34.21$, $p < .001$, $\eta^2 = .426$. This

provides moderate evidence that the findings were not driven primarily by outlier influence.

4.4 Domain-Level Exploratory Analyses

To examine differences across grammatical subdomains, an exploratory MANCOVA was conducted using pre-test scores as a covariate. Given the unidimensional factor structure (58% variance explained), results are interpreted descriptively. Significant differences favoring the

experimental group were found across all domains ($p < .001$). Maximum scores were 6 for subject, predicate, and five nouns, and 12 for indeclinable verbs. The largest effect was observed in the subject domain (partial $\eta^2 = .412$), followed by predicate structures ($\eta^2 = .357$). These findings should be interpreted as exploratory trends rather than evidence of distinct constructs. (see Table 3).

Table 3. Domain-Level Performance by Condition

Grammatical Domain	Experimental M (SD)	Control M (SD)	Partial η^2
Subject (mubtada')	5.1 (0.9)	3.4 (1.1)	.412
Predicate (khabar)	4.6 (1.1)	3.2 (1.0)	.357
The five nouns	4.8 (1.0)	3.0 (1.2)	.398
Indeclinable verbs	9.2 (1.8)	6.8 (1.9)	.389

4.5 Exploratory Productive Grammatical Performance

The productive subscale consisted of four items (three constructed-response items and one error-analysis task) and demonstrated acceptable internal consistency ($\alpha = .81$). Responses were evaluated independently by two raters, yielding a high level of inter-rater agreement ($ICC = .87$). Because the study was powered only to detect effects for the primary receptive outcome measure, no inferential statistical analyses were conducted for the productive subscale. Accordingly, productive performance is reported using descriptive statistics only. In comparison to the control group (Mean = 1.9, Standard Deviation = 0.9), the experimental group produced more items (Mean = 2.8, Standard Deviation = 1.0) than the control group. However, both groups had a lower level

of productive output than the receptive recognition scores measured for each condition (Experimental Receptive Mean = 18.7; Control Receptive Mean = 12.4), indicating that it is likely that participants did not reach full proceduralization of the grammatical knowledge presented over the course of this twelve-hour intervention period. Therefore, these results are best viewed as hypotheses generating data for further research using adequate sample size for productive assessment.

5. Discussion

The results of this pilot study suggest that a V.A.R.K.-based instructional approach improves immediate receptive grammar understanding among eighth-grade students in Jordan. Post-test scores were significantly higher in the experimental group compared to the traditional group, with a large effect size ($\eta^2 = .429$).

Table 4. Summary of Primary and Robustness Findings

Analysis	Statistic	Value	95% CI	Interpretation
ANCOVA (primary)	F (1,51)	38.32***	—	Significant group difference
Effect size	η^2	.429	[.312, .511]	Large, stable effect
Effect size (small-sample)	Hedges' g	1.84	[1.14, 2.54]	Very large effect
Model fit	Adjusted R^2	.415	—	41.5% variance explained
Sensitivity analysis	η^2	.426	—	Stable after outlier removal

These findings indicate that structured multimodal instruction may facilitate more efficient cognitive processing of abstract grammatical structures in early language learning contexts. Given the quasi-experimental design, results should be interpreted as exploratory and associative rather than definitive (Suzuki et al., 2022). The positive effects of the V.A.R.K.-based instructional approach may be explained by the V.A.R.K. model, Cognitive Load Theory, and Skill Acquisition Theory, which together support structured learning through visual, auditory, read/write, and kinesthetic activities, reduced cognitive load, and progressive development from declarative to procedural knowledge. This suggests that instructional design plays a critical role in shaping how learners internalize grammatical concepts. The results support previous findings on multimodal instruction in grammatical learning (Abdul Kareem & Hassan, 2023; Nurhadi & Hilmi, 2022; Al-Zaatreh & Al-Khawaldeh, 2024),

although earlier studies frequently interpreted the V.A.R.K. model within a learning styles perspective. In contrast, the present study integrates the V.A.R.K. model with Cognitive Load Theory and Skill Acquisition Theory and strengthens prior work through fidelity monitoring (92% reliability), robustness checks, and a pilot design.

Several alternative explanations may account for the findings, including novelty effects, researcher involvement, and unmeasured engagement differences, which may have influenced short-term performance gains. Therefore, results should be interpreted as associations rather than causal proof. In summary, the findings indicate that the V.A.R.K.-based instructional approach may enhance receptive grammar learning, but further research with larger samples, delayed testing, and stronger experimental controls is required for confirmation (see Table 5).

Table 5. Interpretive Boundaries of the Present Pilot Findings

The Present Study Supports	The Present Study Does Not Support	What Remains Unknown
Immediate receptive grammatical recognition gains	Long-term grammatical retention	Whether effects persist beyond 8 weeks
Feasible implementation with fidelity monitoring (92% IRR)	Scalability without researcher support	Whether classroom teachers can implement with equal fidelity
Large effects in a pilot classroom context (Hedges' $g = 1.84$)	Population-level generalizability	Whether findings replicate in other populations
Association between V.A.R.K.-based instruction and higher post-test performance	Definitive causal superiority over alternative instructional approaches	Whether an active control condition would produce similar effects
Preliminary productive-performance patterns	Transfer to spontaneous grammatical production	Whether students apply knowledge in authentic communication

5. Conclusion

As a pilot study with both clearly defined methodological and contextual constraints, this research did not allow for definitive conclusions to be drawn about the effectiveness of using

the V.A.R.K. instructional approach within an educational environment focused upon the grammar of Arabic. Rather than providing support for such definitiveness, this research provided preliminary, hypothesis-generating data that suggested that when utilizing a V.A.R.K.

based instructional model, teachers could implement said models at a very high level of quality (92% inter-rater agreement). Further, the use of this model was found to have a positive association to students' post-test performance on a measure of receptive knowledge related to Arabic grammar ($\eta^2 = .429$; Hedge's $g = 1.84$).

Although the results demonstrated statistically significant differences favoring the treatment group, as well as moderate to large effect sizes that remained stable across sensitivity and bootstrap analyses, the findings were limited to the immediate receptive identification of grammatical concepts. Consequently, the results may not generalize to long-term retention, the ability to produce grammatically correct sentences, or student populations beyond those represented in the current sample, which consisted exclusively of male students from a single school. Although the current study demonstrates positive outcomes of using a V.A.R.K.-based instructional approach to teach Arabic grammar within middle-school classroom settings; however, it is recommended that the results be considered as exploratory and not yet generalizable until they are confirmed by larger scale research. It would be reasonable for teachers to test specific aspects of the V.A.R.K. instructional approach, such as the use of multiple senses to help students remember grammatical rules and concepts, before using them as part of their formal classroom instruction. In addition, these new instructional practices could be used in conjunction with, but not replace existing methods of teaching. It is also recommended that textbook authors, publishers and other curriculum developers refrain from making significant changes to their instructional materials based solely upon this pilot study. Rather, if subsequent studies confirm the effectiveness of the V.A.R.K. instructional approach, then there will be greater confidence in recommending the use of V.A.R.K.-based instructional practices throughout an entire school district or state. Professional development programs may introduce the V.A.R.K. instructional approach as an

additional instructional framework for educators. These introductory modules can focus on the practical application of the V.A.R.K. instructional approach to support grammar instruction in Arabic language education, while acknowledging that the current evidence base remains preliminary and requires further validation through larger-scale studies.

A priority for future studies should be multi-site randomized controlled trials (RCTs), conducted in a variety of educational environments in Jordan, including schools in rural areas, private schools, and varying socio-economic backgrounds. Another important area of future research should be longer-term follow-up data collection after the initial 8–12-week period used in this study. Future studies need to collect data at regular intervals over a period of several months or years to determine whether the improvements in learning observed during the initial 8–12-week period are retained over time. Additionally, researchers conducting future studies need to ensure that the control group has equal amounts of instructional time, attention, and engagement as the experimental group. Researchers also need to ensure that they have an adequate sample size and establish inter-rater reliability for their assessments of productive grammar. Finally, researchers need to examine whether instruction based on the V.A.R.K. model reduces cognitive load for learners and increases proceduralization (the process of internalizing knowledge into automatic memory).

This pilot study demonstrated that a V.A.R.K.-based instructional approach may support receptive grammatical learning within a controlled middle-school educational context. Therefore, pending confirmation of this finding through larger-scale research utilizing delayed outcome measures and productive grammar assessment instruments, the results of this pilot study represent a conceptual and methodological starting point for developing future research on V.A.R.K.-based instructional practices and Arabic language education in Jordan.

6. References

- Abdul Kareem, D., & Hassan, Z. (2023). The effect of the V.A.R.K. learning style model on acquiring grammatical concepts among fifth-grade primary students. *Journal of the College of Basic Education, Mustansiriyah University*, 29(145), 165–186. <https://doi.org/10.35950/cbej.v29i120.10690>
- Al-Hawamdeh, M. F., & Ashour, R. Q. (2015). The efficiency of using the learning cycle strategy in acquiring Arabic grammatical concepts: A study applied to grade-nine students in Jordan. *Jurnal Pendidikan Malaysia*, 40(1), 49–55. <https://doi.org/10.17576/JPEN-2015-4001-07>
- Almelhes, S. A. (2024). Gamification for teaching the Arabic language to non-native speakers: A systematic literature review. *Frontiers in Education*, 9, 1371955. <https://doi.org/10.3389/educ.2024.1371955>
- Al-Shammari, A. (2020). The effect of the K.U.D strategy on acquiring grammatical concepts among first grade intermediate students. *Journal of Educational and Psychological Research*, 17(67), 525–547. <https://search.shamaa.org/fullrecord?ID=312574>
- Al-Zaatreh, A. S., & Al-Khawaldeh, M. A. (2024). The effect of instructional scaffolders on the acquisition of syntactic concepts among eighth grade students in Jordan. *Dirasat: Educational Sciences*, 51(2), 149–163. <https://doi.org/10.35516/edu.v51i2.6369>
- Brosh, H. (2017). Grammar in the Arabic language classroom: Perceptions and preferences. *Al-'Arabiyya*, 50, 25–52. <https://www.jstor.org/stable/26451395>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge. <https://doi.org/10.4324/9780203771587>
- DeKeyser, R. M. (2003). Implicit and explicit learning. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 313–348). Blackwell Publishing. <https://doi.org/10.1002/9780470756492.ch11>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE Publications. <https://books.google.jo/books?id=QddDEAAQBAJ&printsec=frontcover#v=onepage&q&f=false>
- Ellis, R. (2006). Current issues in the teaching of grammar: An SLA perspective. *TESOL Quarterly*, 40(1), 83–107. <https://doi.org/10.2307/40264512>
- Filippou, K., Acquah, E. O., & Bengs, A. (2025). Inclusive policies and practices in higher education: A systematic literature review. *Review of Education*, 13(1), e70034. <https://doi.org/10.1002/rev3.70034>
- Hair, J. F., Babin, B. J., Anderson, R. E., & Black, W. C. (2019). *Multivariate data analysis* (8th ed.). Cengage Learning. <https://www.cengage.uk/c/multivariate-data-analysis-8e-hair-babin-anderson-black/9781473756540>
- Hasan, A. (2011). *Al-Nahw al-wafi* (Vol. 1). Dar Al Maaref. https://books.google.jo/books/about/al_Nahw_al_wafi.html?id=TyOdzgEACAAJ&redir_esc=y
- Khasawneh, N. A. S. (2022). The effectiveness of a flipped classroom strategy in developing grammatical concepts among fifth grade primary school students. *European Journal of Educational Research*, 11(1), 207–216. <https://doi.org/10.12973/eu-jer.11.1.207>
- Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge University Press. <https://doi.org/10.1017/9781316941355>
- Mekki, T. A. (2023). The use of VARK in teaching idioms to Iraqi EFL students. *Alustath Journal for Human and Social Sciences*, 62(1), Article 26. <https://doi.org/10.36473/ujhss.v62i1.2005>
- Ministry of Education, Jordan. (2019). *National report on students' achievement in Arabic language education*. Amman, Jordan: Author.
- Ministry of Education, Jordan. (2020). *Jordanian national curriculum framework for Arabic language*. Amman, Jordan: Author.
- Ministry of Education, Jordan. (2022). *National assessment report of Arabic language achievement for basic education students*. Amman, Jordan: Author.

- Nancekivell, S. E., Shah, P., & Gelman, S. A. (2020). Maybe they're born with it, or maybe it's experience: Toward a deeper understanding of the learning styles myth. *Journal of Educational Psychology*, 112(2), 221–235. <https://doi.org/10.1037/edu0000379>
- Nurhadi, N., & Hilmi, D. (2022). Reform for the development of digital Arabic language teaching materials based on constructivism theory. *Izdihar: Journal of Arabic Language Teaching, Linguistics, and Literature*, 5(3), 285–300. <https://doi.org/10.22219/jiz.v5i3.22872>
- Suzuki, Y., DeKeyser, R., & Loewen, S. (2022). The role of practice in instructed second language learning. *Language Teaching*, 55(3), 375–399. <https://doi.org/10.1017/S0261444821000257>
- Sweller, J. (1988). Cognitive load during problem solving: A review. *Cognitive Science*, 12(2), 257–285. https://doi.org/10.1207/s15516709cog1202_4
- Wahba, K. M., Taha, Z. A., & Giolfo, M. E. B. (Eds.). (2023). *Teaching and learning Arabic grammar: Theory, practice, and research*. Routledge. <https://doi.org/10.4324/9781003034209>
- Yu, S., Zhang, E. D., & Liu, C. (2026). Research into practice: Digital multimodal composition in second language writing. *Language Teaching*, 59(2), 197–213. <https://doi.org/10.1017/S0261444824000375>
- Willingham, D. T., Hughes, E. M., & Dobolyi, D. G. (2015). The scientific status of learning styles theories. *Teaching of Psychology*, 42(3), 266–271. <https://doi.org/10.1177/0098628315589505>
- Zhang, P., & Zhang, S. (2024). Multimedia enhanced vocabulary learning: The role of input condition and learner-related factors. *System*, 122, 103275. <https://doi.org/10.1016/j.system.2024.103275>

8. Author Declaration

The authors declare that this manuscript is original, has not been published previously, and is not under consideration for publication elsewhere. The authors further declare that there are no conflicts of interest related to this work. All authors have reviewed and approved the final version of the manuscript and agree to its submission for publication.