

What Works Clearinghouse



Vocabulary Improvement Program for English Language Learners and Their Classmates

Program description	The <i>Vocabulary Improvement Program for English Language Learners and Their Classmates (VIP)</i> is a vocabulary development curriculum for English language learners and native English speakers (grades 4–6). The 15-week program includes 30–45	minute whole class and small group activities, which aim to increase students’ understanding of target vocabulary words included in a weekly reading assignment.
Research	One study of <i>VIP</i> met the What Works Clearinghouse (WWC) evidence standards with reservations. It included 142 English	language learner students in the fifth grade in 16 classrooms in California, Virginia, and Massachusetts. ¹
Effectiveness	<i>VIP</i> was found to have potentially positive effects on reading achievement and English language development.	

	Reading achievement	Mathematics achievement	English language development
Rating of effectiveness	Potentially positive	Not reported	Potentially positive
Improvement index²	Average: +19 percentile points	Not reported	Average: +17 percentile points Range: –4 to +35 percentile points

1. The evidence presented in this report is based on available research. Findings and conclusions may change as new research becomes available.
 2. These numbers show the average and range of improvement indices for all findings across the study.

Additional program information

Developer and contact

Developed by Diane August, Maria S. Carlo, Barry McLaughlin, and Catherine Snow, and now published by Brookes Publishing. Address: Customer Service Department, Brookes Publishing Co., P.O. Box 10624, Baltimore, MD 21285-0624. Email: custserv@brookespublishing.com. Web: <http://www.brookespublishing.com/store/books/lively-6342/index.htm>. Telephone: (800) 638-3775.³

Scope of use

Developed between 1997 and 2000, *VIP* has been implemented in California, Virginia, and Massachusetts. The curriculum is designed for English language learners and native English speakers. Information is not available on the number or demographics of students, schools, or districts using this intervention.

Teaching

VIP is a 15-week program that includes vocabulary activities and related lessons. The program stresses targeted words from a weekly reading assignment. On Mondays English language

learner students are given the weekly reading assignment in their native language to preview before it is introduced in English on the following day. On Tuesdays the teacher leads whole-group lessons to review the text and define the target vocabulary. On Wednesdays teachers divide the students into heterogeneous language groups to complete two cloze⁴ activities. On Thursdays teachers again divide the students into small groups to complete word association, synonym/antonym, and semantic feature analysis activities. Then, each Friday, teachers lead activities that cover a range of topics including analysis of root words and knowledge of multiple meanings of words. The curriculum includes detailed lesson plans, quasi-scripted lesson guides, overhead transparencies, worksheets, homework assignments, and all necessary reading assignment texts.

Cost

The cost of the system is \$39.95 for each of three volumes of the curriculum for students in grades 4, 5, and 6 (one volume for each grade), or \$99 for the full set.

Research

One study (Carlo, August, McLaughlin, Snow, Dressler, Lippman, Lively, & White, 2004)⁵ reviewed by the WWC investigated the effects of *VIP*. It was a randomized controlled trial that met WWC evidence standards with reservations due to differential attrition.⁶

In this study, 17 classrooms from three states were randomly assigned, within each state, to the intervention (n=10) or comparison group (n=7), but one comparison group teacher withdrew from the study prior to the start of the intervention. The 16 classrooms included both fluent English speakers and English

language learners. The study author reported findings for the full sample,⁷ but provided the WWC with data on the English language learner subsample that is the focus of this WWC report. The 16 classrooms included 142 fifth grade English language learner students (94 in the intervention group and 48 in the comparison group). Students in the comparison group received their regular classroom instruction. The study took place over two years.

- ³ *VIP* was published by Brookes Publishing Company, Inc. in 2003. The published program is an adaptation of the research version of the curriculum, and is structured slightly differently. The authors of the Brookes version of the curriculum are Theresa Lively, Diane August, Maria Carlo, and Catherine Snow.
- ⁴ Cloze activities use a fill-in-the-blank scheme, where students are presented a sentence/passage with missing words. They are asked to fill in blanks (that is, “cloze”) with words that will yield a sensible passage, and thereby demonstrate a grasp of word meaning.
- ⁵ The study was also published as a research symposium in 2000.
- ⁶ The study met WWC standards with reservations because one classroom in the control group withdrew from the study. The WWC obtained this information through correspondence with the study author.
- ⁷ The study authors’ analyses did not show an interaction between treatment and language status (English language learner versus fluent English speaker).

Effectiveness Findings

The WWC review of English language learners addresses student outcomes in three domains: reading achievement, math achievement, and English language development.

Reading achievement. Carlo et al. (2004) did not present the actual means for English language learners in the article. They appeared in a graph but could not be easily accessed with precision. The lead author provided the WWC with these students' pretest and posttest means and standard deviations on each outcome measure. The WWC analysis of the English language learner subsample found no statistically significant differences on reading achievement (performance on cloze passages), but the effect was large enough to be considered substantively important by WWC standards.

English language development. Carlo et al. (2004) did not present the actual means for English language learners in the article. They appeared in a graph but could not be easily accessed with precision. The lead author provided the WWC with these students' pretest and posttest means and standard deviations on each outcome measure. The WWC analysis of the

English language learner subsample found no statistically significant and no substantively important impacts on the Peabody Picture Vocabulary Test-Revised (PPVT-R) and the Morphology measure. The Word Mastery impact was statistically significant; the multiple word meanings and word association impacts were substantively positive. The average effect across the five measures of the domain indicated potentially positive effects on English language development.

Rating of effectiveness

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative. The rating of effectiveness takes into account four factors: the quality of the research design, the statistical significance of the findings (as calculated by the WWC⁸), the size of the difference between participants in the intervention condition and the comparison condition, and the consistency in findings across studies (see the [WWC Intervention Rating Scheme](#)).

The WWC found *VIP* to have potentially positive effects for reading achievement and English language development.

Improvement index

The WWC computes an improvement index for each individual finding. In addition, within each outcome domain, the WWC computes an average improvement index for each study and an average improvement index across studies (see [Technical Details of WWC-Conducted Computations](#)). The improvement index represents the difference between the percentile rank of the average student in the intervention condition versus the percentile rank of the average student in the comparison condition. Unlike the rating of effectiveness, the improvement index is entirely based on the size of the effect, regardless of the statistical significance of the effect, the study design, or the analysis.

The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results. The improvement index for the single measure of reading achievement is +19 percentile points. The average improvement index for the English language development measures is +17 percentile points, with a range of -4 to +35 percentile points.

Summary

The WWC reviewed one study on *VIP*. This study met WWC standards with reservations due to differential sample attrition between the intervention and comparison groups. No classrooms left the intervention group, but one of seven teachers

8. The level of statistical significance was reported by the study authors, or where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate the statistical significance. In the case of *VIP*, a correction for clustering was needed.

The WWC found VIP to have potentially positive effects for reading achievement and English language development.

in the comparison group withdrew from the study prior to the implementation of the intervention.⁹ The intervention had potentially positive effects in reading achievement and English

language development. The evidence presented in this report is limited and may change as new research emerges.

References

Met WWC standards with reservations

Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. *Reading Research Quarterly, 39*, 188–215.

Additional sources:

McLaughlin, B., August, D., Snow, C., Carlo, M., Dressler, C., White, C., Lively, T., & Lippman, D. (2000, April). *Vocabulary Improvement in English language learners: An Intervention Study*. Symposium presented at the National Clearinghouse for English Language Acquisition, Washington, DC.

For more information about specific studies and WWC calculations, please see the [WWC VIP Technical Appendices](#).

9. The WWC obtained this information from the study authors.

Appendix

Appendix A1 Study characteristics: Carlo, August, McLaughlin, Snow, Dressler, Lippman, Lively, & White, 2004 (randomized controlled trial with differential attrition)

Characteristic	Description
Study citation	Carlo, M. S., August, D., McLaughlin, B., Snow, C. E., Dressler, C., Lippman, D. N., Lively, T. J., & White, C. E. (2004). Closing the gap: Addressing the vocabulary needs of English-language learners in bilingual and mainstream classrooms. <i>Reading Research Quarterly, 39</i> , 188–215.
Participants	One-hundred forty-two English language learners ¹ in the fifth grade participated in this study. Students were recruited from 16 classrooms in California, Virginia, and Massachusetts. Ninety-four English language learner students were in classrooms randomly assigned to the intervention group, and 48 students were in classrooms randomly assigned to the comparison group. Ninety percent (128 students) of the participants had pretest and posttest measures for at least one outcome. Follow-up contact with the first author revealed attrition in the comparison group; one classroom was not included in the analyses because a teacher left the study prior to intervention implementation, but after random assignment of classrooms to conditions (17 classrooms were originally assigned to conditions, but only 16 were in the analysis sample). In addition, some students in the overall sample received a pilot intervention in the fourth grade, and some did not. However, this intervention report focuses on fifth grade outcomes only.
Setting	The California site included classrooms from two schools that primarily served working class Mexican-American children in both bilingual and mainstream classes. The classrooms in Massachusetts were from a school that served working class, mostly Puerto Rican and Dominican students, within both bilingual and mainstream classes taught by bilingual teachers. The Virginia classrooms were recruited from an “English-medium” magnet school that served mainly working class Spanish speakers from the Caribbean and Central America.
Intervention	The intervention implemented in the study was adapted and published by the authors as the <i>Vocabulary Improvement Program for English Language Learners and Their Classmates (VIP)</i> . Students read newspaper articles, diaries, documentaries, and historical and fictional accounts related to the topic of immigration. This 15-week intervention included 30–45 minutes of teaching four days a week and focused on 10–12 target words per week. On Mondays participants were given the weekly text to preview in Spanish. On Tuesdays the text was introduced in English, and target words in the text were discussed. On Wednesdays participants formed heterogeneous groups (based on English language proficiency) and completed two types of cloze activities. On Thursdays participants engaged in word association, synonym/antonym, and semantic feature analysis tasks. Then on Fridays either analysis of root words and derivation, or knowledge of multiple meanings of words was stressed. Three lessons were observed (during weeks 4, 9, and 13), revealing that six of the nine of the intervention group teachers implemented more than 70% of the key lesson elements, two 50%–60%, and one 35%.
Comparison	Students in the comparison group received their regular classroom instruction. The curriculum provided to the comparison group differed greatly across the schools in each region of the country. Teachers in the comparison group received some professional development in vocabulary teaching two years prior to the beginning of the intervention.
Primary outcomes and measurement	The study measured reading achievement using a researcher developed cloze measure. It measured English language development using measures titled Knowledge of Multiple Meanings of Words, Morphology, Word Mastery, Word Association, and the Peabody Picture Vocabulary Test-Revised (PPVT-R) (see Appendix A2 for more detailed descriptions of outcome measures). Assessments were given in the Fall and Spring of the academic year.
Teacher training	Researchers conducted biweekly Teacher Learning Community meetings with intervention group teachers, providing teachers with curriculum materials including detailed lesson plans, quasi-scripted lesson guides, overhead transparencies, worksheets, homework assignments, and all necessary reading materials. At these meetings, researchers facilitated discussions of practices that worked well in previous lessons and aspects of the curriculum that were problematic. The curriculum was not modified as a result of these meetings.

1. Correspondence with the study authors revealed that they did not report treatment effects separately for English language learners and non-English language learners because their analyses did not show an interaction between treatment and language status. The WWC obtained English language learner subsample data from the authors for the purposes of this intervention report.

Appendix A2.1 Outcome measures in the reading achievement domain

Characteristic	Description
Cloze passages	For this researcher-developed measure, students read three stories with six cloze items per story. Each cloze item consists of a sentence with one word deleted; students are to supply the deleted word using contextual information to guide them. Ten of the 18 deleted words were taught during the intervention.

Appendix A2.2 Outcome measures in the English language development domain

Characteristic	Description
Knowledge of multiple meanings of words (polysemy production)	For this researcher-developed cloze measure, students generate as many sentences as possible conveying the different meanings of words with multiple meanings (such as ring and place). Correct responses are scored based on the frequency of the response in the response pool, and each correct response receives one or more points. Common responses receive one point, intermediate responses two points, and infrequent responses three points.
Morphology	For this researcher-developed measure, students are to provide the base form of 27 derived words after being given the derived word and then hearing a sentence with the base form of the derived word omitted. The students are to write the correct form of the target word. Fewer than a third of the words were included in the intervention. Example item: The derived word is read to the participant (“discussion”). Then, the student is provided with a lean sentence context (“What did he want to _____?”) and is asked to provide the word that fits into the sentence (“discuss”—the base word). Other examples of base word-derived pairs include remark-remarkable, nation-national, and migration-migrate.
Word mastery	For this researcher-developed measure, students are presented with 36 target words. Each of the target words is included in the curriculum and followed by four short definitions. Students must select the correct definition from the four definitions. All 36 words were taught during the intervention about two to three weeks before they were tested.
Word association measures (depth of word knowledge)	This task measures the depth of word knowledge by assessing students' knowledge of the relationship between words. Twenty target words each appear in the center of separate pages with six other words around the periphery of the pages. Students must draw a line from the target word to the three most closely related words printed on the periphery. Half of the target words were taught during the intervention.
Peabody Picture Vocabulary Test-Revised (PPVT-R)	For the PPVT-R, students are read a word and then must select the picture related to that word from the four pictures displayed. This is a widely used standardized test.

Appendix A3.1 Summary of study findings included in the rating for the reading achievement domain¹

Outcome measure	Study sample	Sample size (classrooms)	Author's findings from the study		WWC calculations			
			Mean outcome (standard deviation ²)		Mean difference ³ (VIP – comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
			VIP group	Comparison group				
Carlo et al., 2004 (randomized controlled trial with differential attrition)⁷								
Cloze passages	Grade 5	16	2.20 (3.74)	0.28 (4.01)	1.92	0.50	ns	+19
Domain average⁸ for reading achievement						0.50	ns	+19

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. The standard deviation across all students in each group shows how dispersed the students' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between –50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Carlo et al. (2004), a correction for clustering was needed for the finding in the reading achievement domain.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A3.2 Summary of study findings included in the rating for the English language development domain¹

Outcome measure	Study sample	Sample size (classrooms)	Author's findings from the study					
			Mean outcome (standard deviation ²)		WWC calculations			
			VIP group	Comparison group	Mean difference ³ (VIP-comparison)	Effect size ⁴	Statistical significance ⁵ (at $\alpha = 0.05$)	Improvement index ⁶
Carlo et al., 2004 (randomized controlled trial with differential attrition)⁷								
Knowledge of multiple meanings of words (polysemy production)	Grade 5	16	2.38 (3.20)	0.60 (2.51)	1.78	0.59	ns	+22
Morphology	Grade 5	16	16.36 (29.05)	10.93 (30.20)	5.43	0.18	ns	+7
Word mastery	Grade 5	16	8.76 (6.78)	2.24 (5.15)	0.40	1.03	Statistically significant	+35
Word association measures	Grade 5	16	4.70 (6.75)	1.55 (7.74)	3.15	0.44	ns	+17
Peabody Picture Vocabulary Test-Revised (PPVT-R)	Grade 5	16	15.13 (21.54)	17.48 (20.86)	-2.35	-0.11	ns	-4
Domain average⁸ for English language development						0.43	ns	+17

ns = not statistically significant

1. This appendix reports findings considered for the effectiveness rating and the improvement index.
2. The standard deviation across all students in each group shows how dispersed the students' outcomes are: a smaller standard deviation on a given measure would indicate that participants had more similar outcomes.
3. Positive differences and effect sizes favor the intervention group; negative differences and effect sizes favor the comparison group.
4. For an explanation of the effect size calculation, please see the [Technical Details of WWC-Conducted Computations](#).
5. Statistical significance is the probability that the difference between groups is a result of chance rather than a real difference between the groups.
6. The improvement index represents the difference between the percentile rank of the average student in the intervention condition and that of the average student in the comparison condition. The improvement index can take on values between -50 and +50, with positive numbers denoting favorable results.
7. The level of statistical significance was reported by the study authors or, where necessary, calculated by the WWC to correct for clustering within classrooms or schools and for multiple comparisons. For an explanation about the clustering correction, see the [WWC Tutorial on Mismatch](#). See the [Technical Details of WWC-Conducted Computations](#) for the formulas the WWC used to calculate statistical significance. In the case of Carlo et al. (2004), a correction for clustering was needed for findings in the English language development domain.
8. This row provides the study average, which in this instance is also the domain average. The WWC-computed domain average effect size is a simple average rounded to two decimal places. The domain improvement index is calculated from the average effect size.

Appendix A4.1 VIP rating for the reading achievement domain

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of reading achievement, the WWC rated *VIP* as potentially positive. It did not meet the criteria for *positive effects* because it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *VIP* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. *VIP* had one study that showed substantively important positive effects.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. The one study reviewed did not show any statistically significant or substantively important negative effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

Not met. The one study reviewed met WWC evidence standards with reservations and showed a substantively important, but not statistically significant positive effect.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. The one study reviewed did not demonstrate any statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.

Appendix A4.2 VIP rating for the English language development domain

The WWC rates the effects of an intervention in a given outcome domain as: positive, potentially positive, mixed, no discernible effects, potentially negative, or negative.¹

For the outcome domain of English language development, the WWC rated *VIP* as potentially positive. It did not meet the criteria for *positive effects* because it had only one study. The remaining ratings (mixed effects, no discernible effects, potentially negative effects, and negative effects) were not considered because *VIP* was assigned the highest applicable rating.

Rating received

Potentially positive effects: Evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: At least one study showing a statistically significant or substantively important *positive* effect.

Met. *VIP* had one study that showed substantively important positive effects.

- Criterion 2: No studies showing a statistically significant or substantively important *negative* effect and fewer or the same number of studies showing *indeterminate* effects than showing statistically significant or substantively important *positive* effects.

Met. The one study reviewed did not show any statistically significant or substantively important negative effects.

Other ratings considered

Positive effects: Strong evidence of a positive effect with no overriding contrary evidence.

- Criterion 1: Two or more studies showing statistically significant *positive* effects, at least one of which met WWC evidence standards for a *strong* design.

Not met. The one study reviewed met WWC evidence standards with reservations and showed a substantively important, but not statistically significant positive effect.

- Criterion 2: No studies showing statistically significant or substantively important *negative* effects.

Met. The one study reviewed did not demonstrate any statistically significant or substantively important negative effects.

1. For rating purposes, the WWC considers the statistical significance of individual outcomes and the domain level effect. The WWC also considers the size of the domain level effect for ratings of potentially positive or potentially negative effects. See the [WWC Intervention Rating Scheme](#) for a complete description.