

This grading system is meant to provide a neutral evaluation of a language program's efficacy. Most research evaluation frameworks are based primarily on evaluating the quality of a study. This type of framework de-prioritizes both the research results, and the depth of research completed. In our framework. Programs are evaluated on: their research results, the quality of their studies, the number of studies completed, and the fundamental pedagogical principles behind a program. We believe that this grading system gives a more fair and comprehensive review of a program's level of demonstrated efficacy.

A+	-4 or more studies with control groups that show a mean effect size of .40 or higher on standardized tests
A	-4 or more studies with control groups that show a mean effect size of .40 or higher, but not on standardized tests -Or 2-3 studies with control groups that showed a mean effect size of .40 or higher, on standardized tests -Or a large scale, longitudinal RCT showed a mean effect size of .30 or higher on standardized tests
A-	-One rigorous study that shows a mean effect size of .40 or higher on standardized tests. -Or 2-3 studies that showed a mean effect size of .40 or higher not on standardized tests
B+	-One study showing a mean effect size of .40 or higher, not on standardized tests -Or one low quality study showed a mean effect size of .40 or higher -Or multiple studies showing a mean effect size between .30 and .39, on standardized tests & the program principles are evidence-based
B	-The program principles are evidence-based, but research studies showed a mean effect size below .30
B-	-The program principles are evidence-based, but there are no direct studies with control groups
C+	-The program principles are not well supported by research and a mean effect size was found between .30 and .39
C	-The program principles are not well supported by research and a mean effect size was found between .20 and .29
C-	-The program principles are not well supported by research and the mean effect size found was negligible, but positive.
D	-The program principles are not well supported by research and there are no studies with control groups
D-	-The program showed a mean negative effect size.

The qualitative grading system is still numerically based and awards grades, based on the program's inclusion of the most evidence based instructional principles. For full language programs, marks are awarded for the following criterion: proper scaffolding (IE foundational skills, transition to comprehensive skills), individualized instruction (IE the program individualizes its curriculum to students needs), explicit instruction, phonemic awareness

instruction, phonological instruction, morphological instruction, fluency instruction, comprehension instruction, sight word instruction, and writing/spelling instruction. Programs are then given a grade based on their percentage of inclusion. For example a program that includes all of the above principles would receive a grade of 10/10.