

The ABC's of Junior Achievement's Program Evaluations

The language of program evaluation and assessment can be intimidating, yet crucial for explaining the impact of JA's programs to educators and funders. Here is a basic description of the evaluation process and methods used to evaluate JA programs. This summary provides a basic overview of the phases of evaluation, test instruments, and statistical analyses.

Three Phases of Evaluation

Phase I – Formative Evaluation. The formative phase involves assessing a program while it is being developed and field-tested. Evaluators gather information on how the status of program implementation and what problems have arisen. To do this, evaluators acquire information from major stakeholders – school principals, teachers, students, classroom consultants, and parents. Often, surveys and focus groups are employed to gather this type of information.

Formative evaluation findings are used to fine tune and, if necessary, revise a program.

Phase II - Summative Evaluation. This phase is outcome-based, and examines the impact of a JA program on student learning. Evaluators measure the knowledge gained by JA students and the critical thinking skills they have acquired, as well as changes in attitudes and behavior. Often, surveys, open-ended questionnaires, and classroom observations, are used to gather this type of information.

Phase III – Longitudinal Evaluation. A longitudinal evaluation tracks the cumulative effect of a JA program on student learning and development. This evaluation can begin at any grade or program level (each grade or age level is called a “cohort”). A longitudinal evaluation can be designed to study a group of JA students over several years to determine the cumulative impact of having participated in a JA program or series of programs.

Whereas the *formative* phase provides an early “snapshot” of a new program's progress, a *summative* evaluation demonstrates its long-term learning effects. In turn, a *longitudinal* study reveals the cumulative impact on student knowledge and behavior from participating in Junior Achievement over several years.

Types of Measures Used

A credible evaluation uses *multiple assessment methods* and *multiple sources* for program evaluation. In the formative phase, evaluators use questionnaires,

on-site school visits, classroom observations, interviews, and focus groups with teachers, students, parents, volunteers, and school principals.

Summative and longitudinal evaluations utilize two additional types of assessment:

Objective-Referenced Tests (ORTs) use multiple-choice questions to measure how well students understand basic program concepts. An ORT typically covers all of the learning objectives for a program, and provides a broad understanding of program impact.

Alternative Assessment Tests measure students' ability to apply concepts, solve problems, make decisions, and use related analytical skills. This type of assessment provides a more in-depth, assessment of learning.

Evaluators may also employ attitude surveys and questionnaires to study program impact.

Analyzing the Results

Researchers use two primary statistical tools to determine the educational value or effectiveness of JA programs.

Statistical significance is a probability statement that indicates if the difference in test scores between groups of students is a real difference or merely a random fluctuation due to chance. For example a *p-value (probability value)* of .001 ($p < .01$) means the statistical likelihood of a particular difference occurring by chance alone is one in a hundred. A *p-value* of .0001 means the odds are one in ten thousand that the difference in scores is due to chance.

A *p-value* of .05 (or one in 20) is generally substantial enough to be considered statistically significant.

Effect Size measures the magnitude of the difference between two groups. The larger the magnitude, the greater the impact on learning. This statistic tells us if the difference in test scores is large enough to be meaningful according to accepted standards in a given field of inquiry (e.g., in educational research, psychological research, etc.). Effect sizes of .30 (E.S. = .30) or higher are often considered educationally meaningful, while an E.S. of .50 is robust, and an E.S. over .75 is considered to be a very strong indication of program effectiveness.

Evaluation provides objective measures of the educational value of JA programs. Equally important, JA's large-scale program evaluations are conducted by external, independent firms. Evaluation results can yield important information used to improve programs and demonstrate their effectiveness to important stakeholders.