

Overview of the WBIEG Level-2 Evaluation Toolkit

A. What is a Level-2 Evaluation?

The term “Level-2 evaluation” comes from Donald Kirkpatrick’s four Levels of evaluation, where the second level **assesses what and how much participants in a course learn** from it.¹

The Evaluation Group of the World Bank Institute has measured participant learning by testing course takers at the very beginning of the course, testing them on an equivalent test at the end of the course, and computing the “**learning gain**” by deducting the average class pre-test score from the average class post-test score.

The evaluation aims to assess the **course effectiveness in imparting knowledge to its participants in order to improve the course**. It does not aim to test individual course takers for accreditation.

B. What is the Level-2 Evaluation Toolkit?

The Toolkit **helps course organizers develop, administer and analyze tests that measure what and how much participants learned** in their course. It is composed of **practical** sets of guidelines, templates, macros and databases. Its step-by-step approach focuses on what to do. To keep it short, it includes no theoretical background. It is designed for course providers who want to determine what their participants learned in a simple way, without becoming experts in measuring learning. It was developed for the courses organized by the World Bank Institute. However, other organizations may find it useful. It requires having access to the Internet, and basic knowledge of Microsoft Excel and a word processor.

The Toolkit presents one way to test participant learning. This method was chosen for its simplicity and practicality within the context of the World Bank Institute course. We do not consider it to be the only—neither the best—way to measure participant learning, but it has proven to be effective, when money and time are limited. We do not recommend using this method for accreditation without involving professional experts in test development, learning assessment, or psychometric analysis.

C. What course is a good candidate for a Level-2 evaluation?

Although the Toolkit tries to make Level-2 evaluations easy and affordable, it still requires about one week of course content experts and one week of assistants. Therefore, we recommend that a course team reserve the evaluation for important courses that:

- include **learning** as the main objective;
- have their learning objectives **clearly defined before** the course;
- provide **the same** contents to **all** participants in the course;
- have a **large number of participants** (about 30 or more recommended);²
- plan to be **replicated** (possibly with some adjustments) **several times**; and
- have a duration equivalent to **at least five training days**.

¹ Kirkpatrick, Donald L., *Evaluating Training Programs, The Four Levels*, 2nd edition, Berrett-Koehler Publishers, Inc., 1998.

² For technical reasons, the Toolkit can compute the test results of only up to 254 respondents.

D. What are the steps for conducting a Level-2 evaluation?

The steps to conduct a Level-2 evaluation are schematically represented below. To open the relevant guidelines/tools, click the underlined text. Tasks for content experts/course managers are followed by [C], tasks for assistants by [A].

[Toolkit Glossary](#)

[Frequently Asked Questions](#)

