

How to pilot test items (For course content experts/test manager with assistant)

1. **Testing your items before** using them on the actual course participants is recommended. If participants in previous offerings took the test (or part of the test), their responses should be used as part of the pilot test. In addition, pilot tests can be organized with volunteer test takers. Volunteers can pilot the test at different times and in different settings.
2. Pilot test forms should include:
 - the test taker's name (optional)
 - a question on how long the experience of the test taker is in the field of the course
 - the test items (or part of them) **without** the keys¹
 - space below each item to indicate:
 - a. what (if anything) in the item wording tipped them to what they think is the key
 - b. what (if anything) made them hesitate
 - c. any suggestion to rephrase the item.
3. Ask the volunteers to indicate what they think is the key for each item. If they don't know, they should choose the "don't know" option. They should write their comments below each item.
4. Once volunteers took the test and gave their first comments, give them the keys. Ask them to indicate if they feel that the key is inaccurate and, if so why.
5. Collect the test forms, review the comments, and score the tests. Interpreting the results of the pilot test depends on who answered as explained in the table below.

Who can be volunteer test takers	How to interpret their results
Other content experts: colleagues, peers, presenters of other sessions in the course can answer items that they did not write.	Content experts are expected to score high and to choose the same answers. If experts disagree on a key, revise or replace the item.
Alumni: Participants in former offerings of the course. They may be gathered for a follow-up meeting (e.g., on the implementation of action plans developed at the former offering) and take the test (if they have not already answered the same items during the course).	Alumni are expected to score generally higher than the participant look-alikes and the non-content experts, but lower than the content experts. Alumni's scores give a sense of the test difficulty, in that they are likely to be higher than the participants' scores on the pre-test and lower than the participants' scores on the post-test, assuming that the course targets a group of participants similar to the alumni.
Participant look-alikes: people with about the same level of course-related knowledge as the participants. They can be gathered in Brown Bag Lunches (BBL) or other presentations and take the test (or part of the test split among the BBL participants). Participant look-alikes should not be the actual participants, not even participants in a subsequent course offering.	Their scores are expected to increase with their years of experience. On average, participant look-alikes' scores are expected to be between the alumni's and the non-content experts' scores.
Non-content experts: junior colleagues, assistants, peers who are expert in other fields	Non-content experts are expected to score low. If they score over 50%, make easy items harder.

¹ You may split the test items among volunteers, provided that all test items get piloted. Asking people to answer both items in each pair helps to see if each item is as difficult as its paired item. If both items that form a pair are asked on the same test form, don't make them follow each other.