Using OpenLabyrinth for Script Concordance Testing (SCT)

## Background

Some basic principles about SCT:

1. Script Concordance Testing is best used for testing clinical reasoning. There is much more literature about it in general psychology, especially in medical education assessment (1–3).

2. SCT is somewhat similar to Situational Judgment Testing (SJT) in a few ways:

 a: it tries to avoid the ‘single best answer’ concept as to which response is correct.

 b: the correct response is determined by a reference panel of experts, not by what the question author considers to be correct. i.e. a baseline reference panel set of answers is needed for the set of questions.

## How SCT scoring works

This is better described elsewhere(1). There are some simple points to convey here:

1. It avoids single best answer. So an answer is not just correct/incorrect as dictated by the case author.
2. You get maximum marks (usually 1.0) if your response agrees with the most common response given by the reference group (their modal response).
3. You get no marks if none of the reference group selected the same response as you.
4. You score a proportion of that possible mark depending on how popular your response was with the reference group. So if half as many referees agreed with your response compared with the most popular response, you will get 50% or 0.5
5. It is easiest to work through a table of examples to understand how this works (4).
6. In a well-designed SCT question, there should be a reasonable spread of responses from the reference group. If they all agree on a single response, then this is just the same as a single-best-answer.
7. We explore the spread or peakiness of this response curve further down this document (see below under ‘How good is the SCT approach?’).

## How to set up an SCT Question

For the question stem and context, you can either place this into the general Node text or you can place it into the stem for the SCT QUestion. It will work either way but we generally find it easier to place the stem into the Node’s text because it is then easier to see which Node and question is which, if you have to change the order of your questions. It is also easier to reuse QUestions.

 In Script Concordance Testing (SCT) questions, it is relatively easy to create sets of Questions that can be reused.

There are a few important design points to keep in mind:

1. SCT questions look and act identically to Pick Choice (or radio button) questions. But you must indicate that a Question is an SCT question if it is to be included and scored as such in a SCT Scenario Report. (See below)
2. You can flip a question type back and forth from SCT to Pick Choice, when editing your Question sets.
3. Pick Choice questions can have any number of responses. SCT questions must have 5 responses – this is a requirement of the scoring formulae.

The good news is that if you are using a mixture of SCT and SJT and other questions in your exams, you can still use the same map[[1]](#footnote-1). It does not matter in which order the questions are posed within the exam map. SCT and SJT questions can be freely interspersed or even included on the same Nodes. So for the exam design, it is possible to have a common stem and context, describing a particular patient or situation, and then to ask SCT, SJT or other question styles, all on the same Node or page.

## How to set up and run a Scenario

In a similar manner to how we use Scenarios in Situational Judgment Testing (SJT), you must also use an OpenLabyrinth Scenario in order to produce a table of SCT scores. For more information on Scenarios generally, please consult the main User Guide, either on your own OpenLabyrinth server (see under Help), or [here on the main OpenLabyrinth web site support page](http://openlabyrinth.ca/user-guide/). There is also a small series of [pages that specifically relate to Scenarios](http://openlabyrinth.ca/working-with-openlabyrinth-scenarios/).

There are a few wee glitches which tend to show up when running a Scenario:

1. A user **must access a map via its parent Scenario** if their scores are to be included. Depending on how you have set up the access controls to your map, it may be possible for a user to get directly into the map’s pages without going in via the Scenario – if they do, it is difficult but possible to add them afterwards[[2]](#footnote-2). Make a note of the username and when they made their attempt.
2. When users open the Scenario via the menus Scenarios | My Scenarios, there should be a green [Play] button beside one of the steps. If the Scenario Director has not turned on that Step, it may not be accessible yet.
3. If the Scenario Director has indeed activated that Step but users still cannot see the [Play] button then the Scenario may have been played before using the same username. If so, the Scenario Director will need to Reset the Scenario – see Scenario Manager page. In most situations, you should choose the smaller button, or else you might **lose all of the Scenario scores** from this and previous runs.
4. When the Scenario is running, the Scenario Director can watch their progress on the Manage Scenarios page. You will need to **refresh the page** periodically to see changes. Once they reach the end of the map, usually via the [End session and view report] button, their status will show the green check mark and FINISH status.

## How to set up a SCT Report

Just as with SJT scoring, you will actually need to set up two Scenarios: one to store the scores of your expert panel or reference group; and one to store the scores of your exam participants. The good news is that if you are using a mixture of SCT and SJT and other questions in your exams, you can still use the same two Scenarios: one for the experts and one for the examinees. It does not matter in which order the questions are posed within the exam map. SCT and SJT questions can be freely interspersed or even included on the same Nodes.

So for the exam design, it is possible to have a common stem and context, describing a particular patient or situation, and then to ask SCT, SJT or other question styles, all on the same Node or page.

A Scenario is simply a set of maps and a set of users, all in a single unit. For assigning sets of Users, you can make use of User Groups (see under the Tools menu in your OpenLabyrinth editor). This is a quick way to repeatedly assign the same Group of Users to different Scenarios.

A Scenario can also hold several maps, and then can be grouped in Steps. This allows the Scenario Director more control over when the users can access different parts of the Scenario. So you can, for example, create a first Step with a few practice questions, but not open up the main part of the exam until the Scenario Director is ready to move everyone along to that next Step.

The Scenario Director can choose not to include all users when running a Report of the scores. Simply use the check boxes to show which users should be counted. If this Scenario is to be used as a reference panel, remember to also check at least one of those users as an Expert.

When running a report to generate the scores in either a SCT or SJT exam, open the Scenario and **make sure that some users are selected**. Select the Report Type, which is likely to be the cyan [SCT Report] option, shown below. We apologize for the confusing row of royal blue buttons just above e.g. [Get SCT Report] – these are for use with the to/from time slot options just above. Ignore these for the moment.



Figure : Scenario Manager screen - select which users to include, type of report and reference group

Some exams will allow examinees to go back and **revise** their answers on previous Nodes. (See ‘Allow revisable answers’ on the Details page for each map, for more on how to control this). Choose whether to score their [Last attempt] or their [First attempt] at a question.

Finally, and most importantly, do not forget to choose your **reference panel** from the drop-down list just to the right of the [First attempt] button. This is the Scenario that holds the scores of your expert group. When generating the scoring panel for the reference group itself, choose the same Scenario as the one you are now reporting on. That is, you would make your first expert panel self-referential.

For those who are interested, you can actually compare multiple reference panels with each other to see how this affects the overall scoring, for either SCT or SJT exams. Indeed, as you ran multiple iterations of an exam, you might accumulate a larger set of experts, or eliminate some outliers.

Once you are ready to run your report on the selected candidates, some of which must have finished the map, as shown by a green FINISH check mark, click on the small blue eye beside the map name - . This will generate the SCT report. This will appear in your Downloads folder for your browser when completed.

Warning: note that the math steps required are quite **time consuming** and it will often take 5-10 minutes to generate a report. Depending on the settings on your OpenLabyrinth server, this will sometimes exceed the timeout limits and the report will not complete properly.

Depending on the complexity of the exam and the number of cases or questions, we often found we had to limit the number of candidates being processed to about 20-25 at a time. This is a bit of a nuisance but you can fairly easily combine the generated spreadsheets to bring these multiple iterations together, using Copy & Paste in Microsoft Excel.

This is not an ideal situation. If any group is interested, we can look into optimizing the underlying code so as to make the generation of these reports more time and resource efficient. Please contact us.

## How good is the SCT approach?

In our exams and assessments, we have become increasingly keen on question formats which move away from single-best-answer (SBA). While SBA has the longest track record and is the favorite of psychometricians and item-response theorists as being the most reliable of formats, in real life we are often faced with several possible solutions to a problem from which we have to choose the “least worst” alternative.

In principle, SCT and SJT both use an approach where the correctness of your response is compared to a reference panel of experts or near-peers. This has the advantage that the correctness is determined not by the question author but by an evidence-informed comparison with what others of similar experience might choose.

This raises some interesting questions that are currently being debated in the assessment world. For example, in SCT, the shape of the curve showing the relative preferences of the expert panel matters a lot. In general terms, one is looking for good but not perfect concordance between the experts. After all, if there is perfect concordance and all the experts agree, then this is no different from a SBA style question: only one answer is correct.

And if the peakiness of the curve is too flat, then it matters little which choice is made, which indicates a non-discriminatory question. So what number indicates an acceptable degree of peakiness (kurtosis[[3]](#footnote-3)) or concordance? Remember that these are non-normal curves or Poisson distributions. Indeed, there is an advantage is questions are skewed to left or right; otherwise a canny participant can score more highly by simply picking the middle options.

More importantly, what should be done if there is a saddle-shaped curve or bimodal distribution? Some experts maintain that such a question should be rejected, as it indicates a poorly crafted or ambiguous question. Others hold that this is what life is like and that such situations also arise in real life. This is keenly debated at present. It is up to the exam administrators as to what tack they wish to take in such circumstances.

There are many aspects of SCT that remain worthy of further exploration. We are keen to explore these potential variations with groups who would like to use OpenLabyrinth as an SCT platform.

## References

1. Charlin B, Roy L, Brailovsky C, Goulet F, van der Vleuten C. The Script Concordance test: a tool to assess the reflective clinician. Teach Learn Med. Taylor & Francis; 2000;12(4):189–95.

2. Charlin B, Brailovsky C, Leduc C, Blouin D. The Diagnosis Script Questionnaire: A New Tool to Assess a Specific Dimension of Clinical Competence. Adv Health Sci Educ Theory Pract [Internet]. 1998;3(1):51–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/12386395

3. Goulet F, Jacques A, Gagnon R, Charlin B, Shabah A. Poorly performing physicians: Does the script concordance test detect bad clinical reasoning? J Contin Educ Health Prof. 2010;30(3):161–6.

4. Fournier JP, Demeester A, Charlin B. Script concordance tests: guidelines for construction. BMC Med Inform Decis Mak [Internet]. 2008;8:18. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2427021&tool=pmcentrez&rendertype=abstract

1. *The terminology that we use in OpenLabyrinth for maps and cases can get confusing here. Most of the time, a map (its correct term) describes a single virtual patient case, with many branching choices for that patient. And so, it is often also referred to as a case. However, in SCT and SJT exams, a case is more likely to refer to single patient stem, which may have 2-4 questions arising from it. In this event, a case is more likely to be confined to a single Node.* [↑](#footnote-ref-1)
2. *This requires direct access to the SQL tables on the database. It is even possible to add a whole bunch of users in this manner to generate a post-hoc report, if a Scenario was not set up in advance. Contact us for details on how to do this. It is not for the faint of heart!* [↑](#footnote-ref-2)
3. Kurtosis scores themselves are hard to follow, not aligning with usual variation from 0 to 1. In true terms, they are only an approximation of peakiness. [↑](#footnote-ref-3)