**Data Analysis Guide (to accompany the Speciation Concept Inventory Excel Template file)**

 This guide is intended to help you analyze data that comes from using the Speciation Concept Inventory and is manipulated in the associated Speciation Concept Inventory Excel Template file. On opening the file, you will see that there are **five** worksheets (PRE-test Difficulty, PRE-test Discrimination, POST-test Difficulty, POST-test Discrimination, Learning Gains); these sheets will calculate the difficulty and discrimination indices for each question, as well as the class means, and the all-important student-by-student and whole class learning gains.

**The PRE-test Difficulty and POST-test Difficulty sheets**

* These sheets are for calculating student performance on the Speciation Concept Inventory on an individual question and whole inventory basis for the pre- and post-tests.
* Each student should be given a unique identifier, but this should not be traceable to comply with confidentiality requirements (e.g. not a student ID, cell phone number etc.).
1. In row ‘8’ go to ‘Insert’ -> ‘Rows’. The template file has been set up for just six students but you will need to insert the appropriate number of rows for all of your students.
2. Fill in each row with student answers to the 14 Speciation Concept Inventory questions; input/code ‘1’ for answer A, ‘2’ for answer B, ‘3’ for answer C, ‘4’ for answer D, and ‘5’ for answer E.
3. Check the “SUM TOTAL” row to make sure that the sum totals are the same for all questions (if they are not, you are missing student data somewhere).
4. Check the formulae in the “TOTAL As, TOTAL Bs, TOTAL Cs, TOTAL Ds, and TOTAL Es” rows. As you add rows for each of your students, you will likely need to adapt the **ranges** in the formulae. For example, in the “TOTAL Cs” rows, the formulae for Q1 will look like this: “=(FREQUENCY(B2:B7,3)-(B9+B10))”. You will need to change this to: “=(FREQUENCY(B2:Bxx,3)-(Bxy+Bxz))”, where xx is the final row of student data, xy is the row for “TOTAL As”, and xz is the row for “TOTAL Bs”.
5. You can see the **difficulty index** for each question by looking at the “% As, % Bs, % Cs, % Ds, and % Es” rows. One of these will be the expert answer and show the percentage of your class that answered in the same/desired way.
6. You can also see whether there are any **“Most Common Incorrect Answers” (MCIAs)** present in student thinking by looking at these same rows. Because there are five options (answers A, B, C, D and E) for each Speciation Concept Inventory question, we define an MCIA as a non-expert answer chosen by > 20% of the class (see Smith and Knight 2012). MCIAs suggest a widespread misconception held by your students.

**The PRE-test Discrimination and POST-test Discrimination sheets**

* These sheets are for calculating question discrimination indices for each question of the Speciation Concept Inventory.
1. In row ‘8’ go to ‘Insert’ -> ‘Rows’. The template file has been set up for just six students but you will need to insert the appropriate number of rows for all of your students.
2. Initially, **copy** all the student data for questions 1 to 14 from the appropriate Difficulty worksheet (either the PRE-test or POST-test sheet) before moving back to the appropriate Discrimination worksheet (either the PRE-test or POST-test sheet) -> left-click in cell ‘A2’ -> go to ‘Edit’ -> ‘Paste Special’ -> select ‘Values’ -> ‘OK’.
3. For each question, you will need to score/code the responses as either ‘1’ for expert-like answers or ‘0’ for non-expert-like answers. Initially, check to see whether A, B, C, D, or E is the expert-like answer for the question you are scoring. Recall that ‘1’ = A, ‘2’ =B, ‘3’ = C, ‘4’ = D, and ‘5’ = E. For each question, select all student data in the column that relates to the question you are scoring and go to -> ‘Edit’ -> ‘Replace’ (replace the non-expert answers with ‘0’, replace the expert answer with ‘1’).
4. When you have scored all questions, go to column P (“TOTAL SCORE”) -> right-click in cell ‘P7’ -> Copy -> Select all rows below this with student data in adjacent columns-> ‘Paste’.
5. Repeat step 4 for column Q.
6. Check that the “MEANS” formulae are taking averages for all student data (you might need to input a new range, as currently the formula is set to: “=AVERAGE(P2:P7)”.
7. You must now sort the students based on their overall scores across the whole 14-question Speciation Concept Inventory. Select all student data (row 2 to row xx, from column A to column Q) -> ‘Data’ -> ‘Sort’ -> ‘Sort by TOTAL SCORE (or column P)’, ‘Order Largest to Smallest’.
8. You are going to divide students into equal thirds (Top, Middle, Bottom) based on their overall scores. Initially, input the total number of students in the “Total N” row (column B to column O). Then, divide the total number of students by 3 to work out the number of students that must appear in the Top and Bottom groups.
9. You will need to alter the ranges in the formulae in rows “Top 33%”, and “Bottom 33%”. Now, the formula for “Top 33%” for Q1 (Column B) reads: “=SUM(B2:B3)”. You will need to change this to: “=SUM(B2:Bxx)”, where xx is the final row of the Top third of your students. Do the same for the ‘Bottom 33%” formula.
10. You will need to alter the formulae in row “DI”. Now, the formula for “DI” for Q1 (column B) reads: “=(B11-B12)/2”. The B11 and B12 should update automatically to reflect the cells containing data for “Top 33%” and “Bottom 33%” for Q1, but the second part will not. You should change the formula to: “=(Bxy-Bxz)/Nt”, where xy and xz are the automatically updated cell numbers, and Nt is the total number of students / 3.
11. Having done this for Q1, you need to do it for Q2 to Q14. Simply, select the four cells of column B data that comprise: “Top 33%”, “Bottom 33%”, “Total N”, and “DI” -> ‘Edit’ -> ‘Copy’ -> Select the cells alongside these for Q2 to Q14 (columns C to O) -> ‘Paste’.
12. You can now see the Discrimination Index for Q1 to Q14. Discrimination values of ≥ 0.2 are considered good, and those of ≥ 0.4 are excellent (Morrow et al. 2005, Streveler et al. 2011). Good discrimination values indicate that students who scored relatively well on the full Speciation Concept Inventory also outperformed relatively low scorers on the individual question you are analyzing. Good discrimination values suggest a question is a good guide as to whether students are likely to score well on the whole inventory, and, thus, whether their conceptual knowledge is good.

**The Learning Gains sheet**

* This sheet is for calculating the learning gains of your students on an individual and whole-class basis
1. In row ‘8’ go to ‘Insert’ -> ‘Rows’. The template file has been set up for just six students but you will need to insert the appropriate number of rows for all of your students.
2. Open a new worksheet (you are not going to save this, but need it to match students easily for the next steps).
3. Initially, go back to the ‘PRE-test Difficulty’ sheet and copy all the student personal identifiers (column A) and paste these into your new sheet.
4. Go back once more to the ‘PRE-test Difficulty’ sheet and copy all the “TOTAL SCORE” and “TOTAL %” student data. Go back to your new sheet, click in cell B2 and go -> ‘Edit’ -> ‘Paste Special’ -> select ‘Values’ -> ‘OK’.
5. Repeat steps 3 and 4 by taking data from the ‘POST-test Difficulty’ sheet (all the student personal identifiers in column A, and all the “TOTAL SCORE” and “TOTAL %” data). Go back to your new sheet, click in cell D2 and go -> ‘Edit’ -> ‘Paste Special’ -> select ‘Values’ -> ‘OK’.
6. **Now, this step is critical to make sure you match students correctly before performing the final analyses.** In your new sheet, select all data in columns A, B and C (all the PRE-test data) and go -> ‘Data’ -> ‘Sort’ -> ‘Sort by Non-personal student identifier (or column A)’, ‘Order Largest to Smallest’. Now, select all data in columns D, E and F (all the POST-test data) and repeat, by going -> ‘Data’ -> ‘Sort’ -> ‘Sort by Non-personal student identifier (or column A)’, ‘Order Largest to Smallest’. **At this point, you need to make sure that the personal student identifiers in columns A and D match one another. If they don’t, you have an uneven number of students in your pre and post-test data sets, or something else has gone wrong.**
7. You are now going to transfer this data back into the “Learning Gains” sheet. First, select all the newly re-ordered pre-test data (columns A, B and C) in your new sheet and go -> ‘Edit’ -> ‘Copy’ -> click into cell A2 of the Learning Gains sheet and go -> ‘Edit’ -> ‘Paste’. Now, go back to your new sheet and select all the newly re-ordered post-test data minus the personal student identifiers (just select columns E and F) and go -> ‘Edit’ -> ‘Copy’ -> click into cell D2 of the Learning Gains sheet and go -> ‘Edit’ -> ‘Paste’.
8. Click into Cell F7 and go -> ‘Edit’ -> ‘Copy’ -> select all column F cells below with student data in adjacent cells and go -> ‘Edit’ -> ‘Paste’.
9. Select all data from row 2 to the last row of student data, and from column A to column F and go -> ‘Data’ -> ‘Sort’ -> ‘Sort by POST v PRE DIFFERENCE’ (or column F), ‘Order Largest to Smallest’. Depending on the numbers here, you will do one of three things:
10. **IF THE NUMBER IS POSITIVE (the student’s post-test score is higher than their pre-test score):** in cell G2, write the following formula: “=((E2-C2)/(100-C2))”. Then, for all other positive numbers below this in column G, simply copy the formula and paste it by going -> click into cell G2 -> ‘Edit’ -> ‘Copy’ -> select all cells in column G with positive numbers alongside and go -> ‘Edit’ -> ‘Paste’.
11. **IF THE NUMBER IS 0 (there is no difference between the student’s pre-test and post-test score):** simply enter ‘0’ into all column G cells that are alongside these 0s.
12. **IF THE NUMBER IS NEGATIVE (the student’s post-test score is lower than their pre-test score):** in the first row of column G that applies, write the following formula: “=((CXX-EXX)/(CXX))”, where XX is the row number for the student whose score is negative. Then for all other negative numbers below this in column G, simply copy the formula and paste it by going -> click into cell GXX -> ‘Edit’ -> ‘Copy’ -> select all cells in column G with negative numbers alongside and go -> ‘Edit’ -> ‘Paste’.
13. Finally, adjust the range in the formula for the “C-AVERAGE” at the bottom of column G. The formula currently reads: “=AVERAGE(G2:G7)”. You will need to change this to: “=AVERAGE(G2:Gxx)”, where xx is the row number of the last piece of student data.
14. You can now see the learning changes for all students, and the learning change for the class as a whole (the C-average). These are displayed as proportions of the amount of improvement or regression that could possibly have been made at the post-test stage (see Marx and Cummings 2007).

**References**

* Marx, J. D., and Cummings, K. (2007). Normalized change. *American Journal of Physics* **75** (1): 87 – 91.
* Morrow, J., Jackson, A., Disch, J., and Mood, D. (2005). Measurement and Evaluation in human Performance. Champaign, IL: *Human Kinetics*.
* Streveler, R. A., Miller, R. L., Santiago-Roman, A. I., Nelson, M. A., Geist, M. R., and Olds, B. M. (2011). Rigorous Methodology for Concept Inventory Development: Using the ‘Assessment Triangle’ to Develop and test the Thermal and Transport Science Concept Inventory (TTCI)\*. *International Journal of Engineering Education* **27** (5): 968 – 984.
* Smith, M. K., and Knight, J.K. (2012). Using the Genetics Concept Assessment to document persistent conceptual difficulties in undergraduate genetics courses. *Genetics* **191** (1): 21 – 32