**Learning Dissonance – Overcoming Inter-year Comparisons in Educational Experiments**

**Introduction**

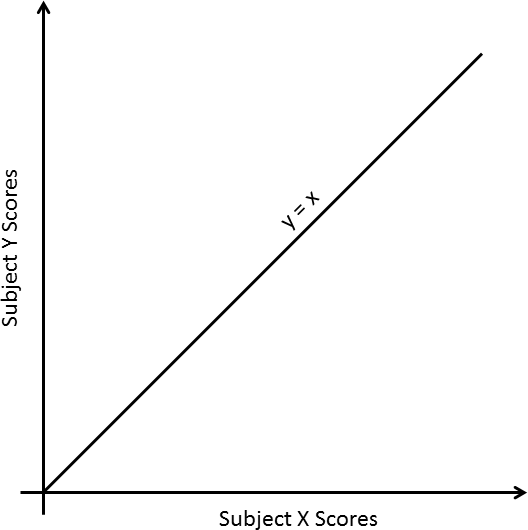
The purpose of the following technique is to overcome the ethical problems when designing experiments in education. If a technique is trialled on two groups of students and the difference between their attainment is measured then one group will advantaged/disadvantaged, which is not good practice.

**Overview**

Assuming that a cohort of students find two courses:-

1. to be equally difficult
2. are assessed to an equal standard
3. are taught to the same standard
4. have an equal access to learning materials.

If these four criteria were to be met then you would expect that the students would get the same results in both courses, i.e. the results from these two courses would form a straight line with the equation y = x, i.e. intercept would be 0 and the gradient would be 1. The correlation coefficient would also be 1.



It is not being assumed that students have:-

* the same starting points for each subject
* the same levels of ability for each subject.

Since the initial four criteria are unlikely to be met, the pairs of scores will form a cloud of points. In other words, the linear correlation coefficient is likely to be <1, whilst the slope and intercept are unlikely to be 0 and 1 respectively.

**The Technique**

If, say, a course is changed in some way then this will alter the pattern of scores for that course. For example, if the standard e-Leaning platform were changed for a course and a new one trialled, then it would be hoped that the scores would improve for that course. However, everyone who has taught knows that cohorts vary. The same lecturer, using the same materials and having the same assessment will get changeable results year on year. The group dynamic will have an effect, there will be an effect from other courses, e.g. another course has several lecturer changes and so a less positive atmosphere is created, as well as changes in the selection criteria compounding the problem.

This means that comparing the average one year to the average the next is problematic. Suppose that the new e-Learning platform works well, but the opinion formers are negative. The overall result may appear unchanged. It will seem as if the new platform has not made any difference, therefore a subtler approach is needed.

It is being argued that instead of comparing year-on-year results for a course, the comparison should be made between the courses of a year. Assuming that there are *n* courses in an academic year, then if the e-Learning platform worked, the trial course results should improve relative to the other courses. It is the relative placing of the courses that needs to be measured and monitored rather than the year-on-year results. The problem then becomes one of measuring these relative changes between courses, a technique we are calling learning dissonance.

There are eight stages needed to complete this analysis. A summary chart of the process is shown at the end.

**Stage 1 - Obtaining the Data**

This will depend upon the system being used. It will usually be necessary to apply for ethical approval if you intend to publish since actual results are needed.

The technique requires the full set of results for all courses (or subjects) and that the student who achieved each result be identified in some way. Their name may not be needed if, say, an enrolment number or other unique identifier can be used. This is useful when seeking ethical approval since it is harder to identify an individual.

**Stage 2 - Creating the Data Files**

Having found the data, it will then be necessary to clean it (extract any missing or anomalous data) and then convert it into either plain text, database or spreadsheet format.

The naming or data design system must make it possible to select courses or subjects which have an academic year and level in common. For example, you may want to select all the courses for 2017/18 and were first years, so naming the text file as, say, *2017-18-1-CourseName.txt* would be helpful.

Similarly, were you to be using a database then having a design which included a table called, say, **Course** with the attributes (CourseID, Year, Level, CourseName) would make analysis easier. The **Result** table would then have *CourseID* as a foreign key.

If a spreadsheet is being used then having a separate sheet for each academic year and level would aid the following calculations. Each sheet would then have the courses or subjects across the top and the students down the side. Cells would be blank where a student did not take a course.

**Stage 3 – Form the Data Cells**

The point of this exercise is to compare like with like. A cell in this system therefore contains all the data sets of results for an academic year and level. This allows the results of individual students in different courses to be compared. The cells are a snapshot of the situation in the institution of courses at a level.

**Stage 4 – Finding Matching Students and Courses**

Matching like with like implies that not all the results of a course will be used. Suppose a course is seen as more difficult, therefore only the more confident take it. The results would be skewed against a course taken by less confident ones. One would expect the results in the first course to be higher than in the second merely because of the type of student taking it.

This technique therefore require that the data sets only includes the results of those students who took both course. For example, there may be 100 taking course A and 80 taking course B. However, this technique requires the selection of students who took both A and B, i.e. the intersection of the two sets, which is likely to be a smaller number, say, 60.

The intersection sets for each pairing of courses within a cell should be created. The results for an example data set are shown below.

**Example Data Set:**

Imagine that a cell contains three courses; A, B and C.

Courses A and B have, say, 52 students in common. The mean score for these 52 students for course A is 59.3. The mean score for the same 52 students for course B is 45.9.

Courses A and C have, say, 21 students in common. The mean score for these 21 students for course A is 67.8. The mean score for the same 21 students for course C is 64.6.

Courses B and C have, say, 67 students in common. The mean score for these 67 students for course B is 62.5. The mean score for the same 67 students for course C is 71.1.

These can be written as letter pairs, for example AB would represent the mean score of group A for those students who also took course B. Therefore, the statements above could be re-expressed as:

AB = 59.3 with n = 52

AC = 67.8 with n = 21

BA = 45.9 with n = 52

BC = 62.5 with n = 67

CA = 64.6 with n = 21

CB = 71.1 with n = 67

**Note**: Courses are not compared with themselves.

**Stage 5 – Calculating the Statistics for a Course**

At this stage, the actual scores of individuals become irrelevant since the data sets contain only students who have taken both courses. Two statistics are needed; the number of students in the intersection data sets and the mean scores for each subject.

For example, let us assume that 52 students take both course A and course B. Course A may have had 97 students on it, but only 52 of those 97 took both A and B. It is only the results of these 52 that are of interest.

The mean score of the 52 students who took course A was, as shown above, was 59.3 whilst those same 52 students scored an average of 45.9 in course B.

The courses now need to be compared. Course A intersects with courses B and C (but some courses could be mutually exclusive and would therefore not be used), but there are 52 in the intersection with course B and only 21 with course C. The result for course C would be overemphasised were a straight average of the means to be taken. This can be overcome by using the weighted averages instead. The weighted mean for each of the courses would be:-

(where x̅ is the mean grade for a course and n is the sample size)

At this stage, we are only interested those rows which start with the course letter.

Course A: (AB and AC)

= [(59.3 × 52) + (67.8 × 21)] ÷ (52 + 21)

= 61.75

Course B: (BA and BC)

= [(45.9 × 52) + (62.5 × 67)] ÷ (52 + 67)

= 55.25

Course C: (CA and CB)

= [(64.6 × 21) + (71.1 × 67)] ÷ (21 + 67)

= 69.55

**Stage 6 – Calculating the Statistics for “Other” Courses**

The weighted means for each course are only part of the answer. The weighted means for the “other” courses also need to be calculated. There were two possible ways to do this; use only those courses that interact with the course being considered or use all courses, i.e. those that do not intersect with the course as well.

The second option was chosen because it represents the placement of a course relative to all other courses with the year. The formula for this would be:

(where !x̅ is the mean grade for all other courses and n is the sample size)

At this stage, we are not interested those rows which start with the course letter.

Course !A: (BA, BC, CA and CB)

= [(45.9 × 52) + (62.5 × 67) + (64.6 × 21) + (71.1 × 67)] ÷ (52 + 67 + 21 + 67)

= 61.33

Courses !B: (AB, AC, CA and CB)

= [(59.3 × 52) + (67.8 × 21) + (64.6 × 21) + (71.1 × 67)] ÷ (52 + 21 + 21 + 67)

= 66.01

Courses !C: (AB, AC, BA and BC)

= [(59.3 × 52) + (67.8 × 21) + (45.9 × 52) + (62.5 × 67)] ÷ (52 + 21 + 52 + 67)

= 57.72

**Stage 7 – Calculating the Positional Score**

The relative positions of the courses can now be calculated. The course weighted average score divided weighted average of the “other” course.

Course A:

= 61.75 ÷ 61.33

= 1.007

Course B:

= 55.25 ÷ 66.01

= 0.837

Course C:

= 69.55 ÷ 57.72

= 1.205

If these were actual figures, they would show that course C performs best relative to the other courses, whilst course B performs worst.

**Stage 8 – Calculating the Learning Dissonance**

The learning dissonance statistic then compares the positional scores between years. Course B seemed to perform badly in the previous stage. This course may the have had an intervention, so the positional scores for the year after may have become:

Course A: 1.001

Course B: 0.937

Course C: 1.155

The interest is then in how much the positional scores have moved. This is the learning dissonance, i.e. the movement in learning between one year and the next. In the example being used the learning dissonance for the courses would be:

Course A: 1.001 – 1.007 = -0.006

Course B: 0.937 – 0.837 = 0.1

Course C: 1.155 – 1.205 = -0.05

**Expected Results**

Three courses do not make a relevant sample. However, if a large enough quantity of learning dissonance scores were calculated then a pattern might emerge. The question then becomes which type of distribution to expect.

The expectation of a teacher/lecturer on entering a classroom at the beginning of a new academic year is that the ability of the students will vary. Both John Carroll   
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
(Carroll 1963) and Benjamin Bloom   
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
(Bloom 1968) argued that these abilities are likely to be normally distributed. To quote Bloom:

“*Each teacher begins a new term (or course) with the expectation that about a third of his students will adequately learn what he has to teach. He expects about a third of his students to fail or to just "get by." Finally, he expects another third to learn a good deal of what he has to teach, but not enough to be regarded as "good students."*”   
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
  
(Bloom 1968)

It therefore seems reasonable to assume, in the absence of evidence to the contrary, that learning dissonance (LD) scores should have a mean of 0 and be normally distributed. If this were, then case then the following would also hold:

1. the probability of changes in teaching methods can be assessed.
2. the effectiveness can be “graded”.
3. the effectiveness of one change can be compared to another.
4. the compound effectiveness of more than one change can be calculated.

**Actual Results**

*(This is likely to be about 1/2 to 1/3 of the paper because of the statistical analysis and commentary.)*

**Alternative Uses**

The discussion so far has been about gauging the effectiveness of changes in teaching methods. These use the course as the unit of aggregation, however, that ned not necessarily be the case. Effects other than teaching methods would also be possible. For example, suppose the quality of teaching facilities varied within an institution. Some buildings might be older or less well maintained, some might be noisier, etc.

The same technique could be used to assess rooming by grouping results by room rather than course. Student results can then be gathered by room and compared to give a positional score for each room. One would expect these positional scores to be more or less constant over time … unless improvements were made to the infrastructure. In this way, the effect of environment could be calculated.

Similarly, the effectiveness of teachers/lecturers could be calculated. This is likely to be a political hot potato, because some might see it as an opportunity to beat those “performing” less well. However, this narrow-minded approach would leave out other abilities that these people bring such as comprehensive materials, being good team players, etc. Instead of grouping by courses, the student results would be grouped by lecturer. Again, assuming little in the way of changes to teaching techniques, the positional scores would remain constant. It could then be ascertained which styles those who performed “better” used and see which, if any, might be used by others. Teaching styles are individual, so what works for one may not work for another.

Similarly, if students were substituted as the aggregation unit for course then changes in student motivation between years could be calculated. Questionnaires or interviews could then be used to see which factors helped or hindered in this area. This might then lead to policy changes.

There are many other factors that could be assessed in this way (are one- or two-hour lectures more effective, etc.). The results from each of these studies could then be modelled to create a local and global (if enough studies were carried out) model of learning. This could then have benefits to both the student and the institution since the use of limited resources (a constant problem in education) could then be optimised.

**References**

Bloom, B.S., 1968. Learning for Mastery. Instruction and Curriculum. Regional Education Laboratory for the Carolinas and Virginia, Topical Papers and Reprints, Number 1. *Evaluation comment*, 1(2), p.n2.

Carroll, J., 1963. A model of school learning. *The Teachers College Record*, 64(8), pp.723–723.

**The Stages Diagrammatically**

