Visualizing data to aid analysis

Introduction

The data that this research project aims to collect will be an extensive blend of quantitative and qualitative metrics – stuff I need to count, and other stuff that has a softer meaning that I will need to think about.

Assuming the deployment of the main research QNR is successful and that respondents reply in the numbers that will be necessary to properly address the underlying research question, I have estimated that I will be faced with sorting and understanding perhaps as many as 100,000 bits of data.

As part of the analysis process, 'radar' plots of the QNR data will be created where each axis of the plot represents each one of the constructs that I am exploring. These constructs are:

- o *Learning Related Emotions*
- o *Anxiety Regulation & Motivation*
- o *Academic Self-efficacy*
- o *Self Esteem*
- o *Learned Helplessness*
- o *Academic Procrastination*

When plotted together, these will visualize a profile of each student respondent.

Recall that the focus of the complete project is first of all, to try to work out if the profiles generated by students who disclose that they have dyslexia are significantly different from the profiles generated by their peers who do not indicate a dyslexic learning difference. From this, it is expected that profiles will emerge from respondents

who are reporting that they do not have dyslexia that are nonetheless more closely aligned with those profiles associated with dyslexic students, and hence reveal students who are showing indications of un-identified dyslexia. In this way, the three, distinct research groups, ND, DI and DNI will have been established.
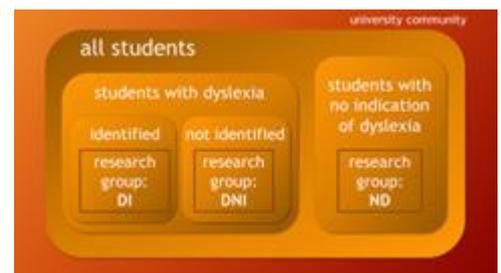
In the second stage of the data analysis the Academic Behavioural Confidence of students in each of the three research groups will be explored from data also collected in the QNR. Recall that the project research question being investigated is whether students with possible, un-identified dyslexia present a higher ABC than their dyslexia-identified peers.

If this proves to be the case, then as outlined in other pages on the project website, this will be a highly interesting result as it will imply that the identification of a student's dyslexia may be a factor in lowering confidence in their academic behaviour. Given the literature supporting linkages between academic confidence and academic achievement, one impact of this result might be that professional staff in the HE sector who take heed of it, could be facing a uneasy dilemma about whether to disclose or not to a student who exhibits a profile aligned with dyslexia – according to the parameters of my research at least – because to do so may then have a significant impact on their potential academic achievement.

Creating radar plots and scatter diagrams using CHART.JS

The application *Chart.js* has been used to generate radar plots to visualize data collected from the preliminary enquiry, DIMENSIONS OF DYSLEXIA, reported on the project webpages (here) and in an earlier StudyBlog post, where a more comprehensive analysis and discussion will follow in due course once the data collected has been more carefully inspected.

This preliminary enquiry has been a useful process as aside from informing the development of the main research QNR, it has enabled the CHART.JS application to be tried out to assess the complexity involved in creating the plots and also to gauge its effectiveness in presenting data. The application has the functionality to create a variety of animated charts based on javascript coding which allows raw data to be entered into chart template scripts. Styles and colours can be easily modified to enable multi-layered charts to be created that are visually attractive but which retain good accessibility through simple formatting. Once an HTML template has been created for the plots, linking the data from the spreadsheet to which it is downloaded from the QNR-generated e-mail has been quite straightforward.

For the DIMENSIONS OF DYSLEXIA enquiry, two chart styles were tested: scatter diagrams and radar plots. The scatter diagrams that are produced by the application are a highly effective visualization of the inter-relationships between each of the 18 dimensions that the QNR collected data about. By pairing every dimension against every other dimension this created 153 scatter diagrams which were attached to a matrix of the corresponding correlation co-efficients – available here. The scatter diagrams easily enabled data outliers to be spotted, which when discounted from the data, allowed more representative correlation coefficients to be re-calculated. Teasing out what all this data means will be the focus of a subsequent StudyBlog post.

Radar plots were also created as a means to view the data collectively. The figure below shows the radar plots for two respondents and in each plot, the grey background plot represents the

mean values for each dimension with the respondents' plot overlaid. (For this preliminary enquiry two sets of plots were constructed, the first showing each respondent's scores dimension by dimension overlaid onto the means, the second set showed all respondents replies to each of the 18 dimensions. The complete sets together with a preliminary analysis is available here).
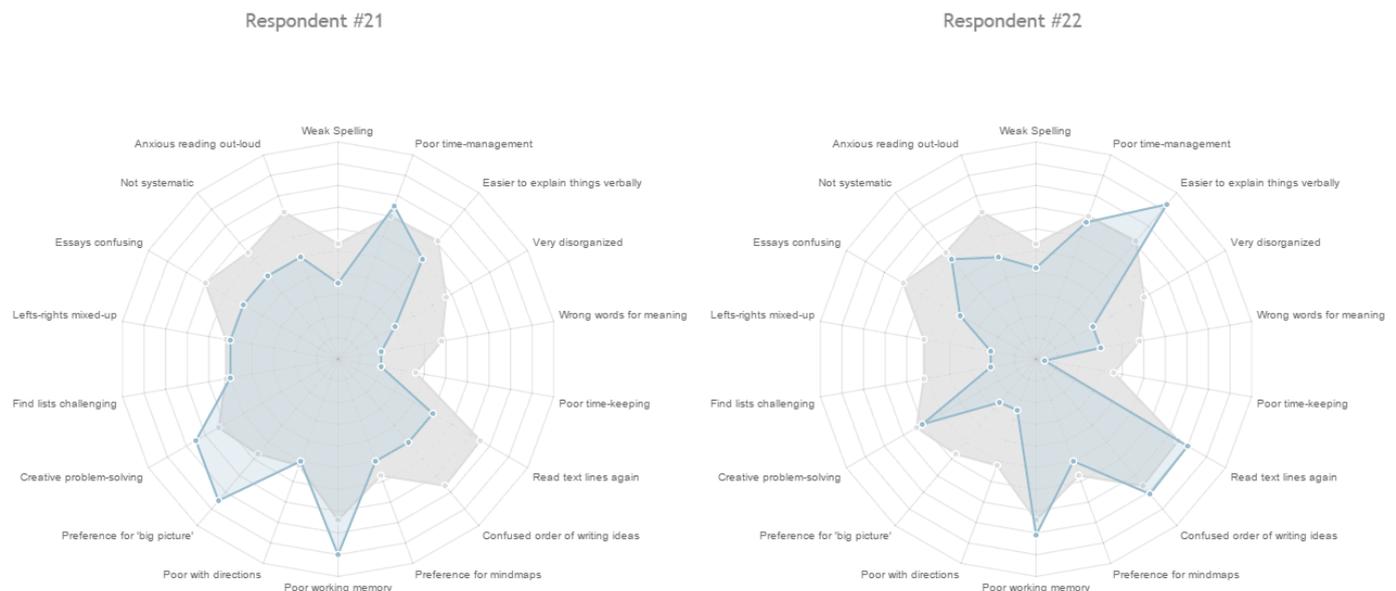
So in advance of deploying the main research QNR shortly (early 2016), it is important to trial how the data collected from this will be visualized in a similar way.

The main research QNR has been extensively tested for functionality and cross-browser compatibility – which has been a lengthy process – and I am now satisfied that it works according to the design brief and that the data generated from the QNR replies can be stored in ways that make it manageable and accessible. However, with a QNR that is comprised of 80 Likert-style stem statements that are grouped into categories that will create 8, distinct Likert scales, developing a way to inspect, digest, analyse and understand all this data is a challenging task. However, I have been fully aware of this and thought carefully about it when scoping out the project design and research methodology.

With this in mind, I wanted to see how I could get a radar plot to best display typical data that will arrive from a QNR respondent. Recall that the intention from the outset, is to build on the 5-axis Locus of Control Profile radar plots that were created to visualize the data collected in the Pilot Study as these proved invaluable in the analysis processes. The complete set of 49 profiles collected in that study is available here.
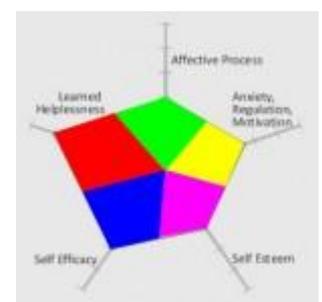
The aim now, is to simultaneously view three plots overlaid onto the same axes: one displaying the means of data from research group DI, those students with known dyslexia; the second displaying the means of data from research group ND, students with no dyslexia; and the third plot being the data from the respondent. In this way, a 3-plot profile will be constructed for each QNR respondent and it is hoped that from these, it will be possible to identify the third research group DNI, that is, students whose profiles are more closely aligned with those of the dyslexia group but who have never been identified as dyslexic. A respondent will have declared in the opening section of the QNR if they have a dyslexic learning difference or not – assuming they are truthful!

There will be six axes to the plot, representing each of the 6 constructs that are being explored in this section of the main research QNR. Five of these are carried forward from the Pilot Study – with AFFECTIVE PROCESS being re-labelled LEARNING RELATED EMOTIONS for this project – and a further construct, ACADEMIC PROCRASTINATION being added. The rationale for this is discussed elsewhere on the project webpages. Each construct is being measured as a Likert scale comprising 6 Likert response item stem-statements. A respondent will have registered their response using the slider control aligned against each stem-statement by adjusting it along its range of 0 to 100, where the zero-end corresponds to 'strongly disagree' with the statement presented, to 100 which will be indicating strong agreement (example below).

The constructs and their stem statements that will be used in the construction of the 3-plot profiles are:

- o Learning Related Emotions
    - o *I am able to settle down to my work anytime, anyplace*
    - o *I feel too embarrassed to ask for help with my studies*
    - o *I feel guilty about my learning challenges*
    - o *I think my student-peers mostly regard my learning challenges as excuses, for laziness for example*
    - o *I don't use any of the learning support services because it makes me feel different*
    - o *I don't think about my learning challenges much*
- o Anxiety, Regulation & Motivation
    - o *I find it quite difficult to concentrate on my work most of the time*
    - o *I don't think my learning challenges make me any more anxious than anyone else*
    - o *I use my learning strengths to help me with study strategies*

- - *I need to work much harder than my friends to get similar grades*
  - *I often feel frustrated when trying to study*
  - *I enjoy my studies even more when the work becomes difficult*
- Academic Self-efficacy
  - *I believe that my learning strengths really make a difference to my academic progress*
  - *I plan and organize my work carefully which I believe helps me to get good grades*
  - *I don't think my learning challenges make any difference to the way I tackle my work*
  - *I approach my written work with a high expectation of success*
  - *I believe my learning strengths help me to be more creative*
  - *I can manage my studies quite adequately without any help*
- Self-esteem
  - *I often felt pretty stupid at school*
  - *If I try hard I can achieve just as much as anyone else*
  - *I think I'm good at studying, perhaps even academically talented sometimes*
  - *I approach my written work with enthusiasm*
  - *At times, I think that I'm hopeless at tackling academic work*
  - *My contributions in class are usually rubbish, so generally I don't bother*
- Learned Helplessness
  - *When I start a new course or topic, I usually think it will be too difficult for me*
  - *I've had help for dealing with my learning challenges but it hasn't made any difference*
  - *I'm generally not surprised when I get a low grade*
  - *I will always be held back by my learning challenges*
  - *I think my grades are as much to do with luck as with any effort on my part*
  - *However hard I try, this rarely makes a difference to my grades*
- Academic Procrastination
  - *I usually finish my essays or assignments well in time for the deadline*
  - *I tend to put off getting started on my essays or assignments until I really have to*
  - *For one reason or another, I often have to request extra time to complete my work*
  - *As soon as I'm given an essay or assignment title, I'm usually eager to get going on it straight away*
  - *My essays or assignments would probably be better if I didn't have to rush to finish them*
  - *I often find other things to do rather than working on my studies*

To generate test data, I completed the QNR myself as honestly as I could by responding to the stem statements exactly as I might if I were working through the QNR on first sight. This provided the 36 raw scores required to begin the construction of the first radar plot.

Recall that the purpose of the plots is to compare the radar plot for the respondent's construct scores against radar plots for the mean construct scores for the research groups DI and ND. However, in generating mean scores, because some stem statements are negatively phrased

whilst others have a more positive sense, an adjustment for this has to be applied before calculating these means. If this were not done, then it is felt that for the constructs' Likert scales which comprised both positive and negative stem statements, a 'false mean' would be the result. In other words, an adjustment needs to be applied to some stem statements in some groups so that all statements in a group exhibit the same polarity for the purposes of calculating mean scores – I have explained this more fully below. Positive-negative phrasing is a common technique in QNR design although usually it is a mechanism for addressing issues of internal reliability such that one attribute/characteristic/dimension is explored through two questions or statements, one using negative phrasing, the other positive. For my questionnaire, rather than using this technique as a means to address internal reliability, I have tried to use a balance of +/- stem statements so as to create neither a wholly negative, nor generally positive sense to the complete questionnaire. I felt that otherwise, invalidities in the data may arise through respondents getting a feeling that all the answers they were providing appeared to be presenting them in a largely negative way and hence either deliberately or indeed subliminally, respond to some statements untruthfully for fear of presenting themselves overall in a bad light, despite the anonymity processes that have been built in to the QNR.

So to deal with this a process has been devised to gain a sense of how workable it will be and to what degree it tries to reduce error and ensure that the data produced is as valid as possible:

Firstly, the 36 stem statements were 'coded' as + or − as a marker of their sense. To do this I asked myself the question: 'would I feel pleased or proud to admit to this [stem statement] or would I feel guilty, shameful or embarrassed?' The former I coded as a +ve stem statement, the latter I coded as -ve.

The table below shows how these adjustments panned out, and also shown are guestimate means for all 36 stem statements as they may have been derived from respondents' QNR replies and adjacent to these, how the +/- adjustment converts them into a consistent construct polarity. Notice that each of the six sections has also been assigned a section bias (polarity), either + or − depending on whether the section has either a majority of positively or of negatively coded stem statements. For example, for the construct LEARNING RELATED EMOTIONS it was felt that 4 of the 6 stem statements portrayed a negative sense against the remaining two being more positive so this construct was assigned a negative bias. So for this construct, the two positively-biased statements are adjusted to align them with their negatively-coded chums. In practice, this means reversing the scores for these two positively-biased statements so that, for example, an ACTUAL (raw) recorded score of say, 70 is reversed into an ADJUSTED score of 30. The columns of adjusted scores are then used to calculate the

construct means. Translating what this means into simple language is that a high score in this construct would be indicating that a respondent is presenting strong negativity in their learning related emotions – guilt, embarrassment, feelings of difference, possibly dwelling on the negative aspects of their learning challenges and so forth. In contrast, for the construct ACADEMIC SELF-EFFICACY all the stem statements are phrased in a positive sense so this construct overall has been assigned a positive bias and no statements' respondent scores need adjusting. This means that a respondent presenting a high score in this section would be demonstrating strong levels of academic self-efficacy – good organizational skills, a good awareness of learning strengths and how these can be positively channelled into their academic progress and achievement and so forth.

ND raw	ND adj	DI raw	DI adj	stmnt
	μ=33		μ=67	+ / -
70	30	40	60	+
50	50	60	60	-
40	40	70	70	-
30	30	70	70	-
20	20	70	70	-
70	30	30	70	+
	μ=70		μ=27	
30	70	70	30	-
70	70	30	30	+
70	70	30	30	+
30	70	70	30	-
30	70	80	20	-
70	70	20	20	+
	μ=70		μ=28	
70	70	20	20	+
60	60	20	20	+
70	70	20	20	+
70	70	20	20	+
70	70	20	20	+
80	80	20	20	+
	μ=72		μ=30	
20	80	70	30	-
80	80	30	30	+
70	70	30	30	+
60	40	30	30	+
20	80	70	30	-
20	80	70	30	-
	μ=20		μ=80	
20	20	80	80	-
20	20	80	80	-
20	20	80	80	-
20	20	80	80	-
20	20	80	80	-
20	20	80	80	-
	μ=33		μ=80	
60	40	20	80	+
30	30	80	80	-
30	30	80	80	-
70	30	20	80	+
60	40	80	80	-
30	30	80	80	-

(research groups ND (no dyslexia) and DI (dyslexia identified); 'raw'/'adj' = mean of raw/adjusted scores; μ = mean for that construct)

Learning Related Emotions [*section bias: - convert +ve statement scores using Adjusted Score = 1 – Actual Score*]
2.1-1   I am able to settle down to my work anytime, anyplace
2.1-2   I feel too embarrassed to ask for help with my studies
2.1-3   I feel guilty about my learning challenges
2.1-4   I think my student-peers mostly regard my learning challenges as excuses; for laziness for example
2.1-5   I don't use any of the learning support services because it makes me feel different
2.1-6   I don't think about my learning challenges much

Anxiety, Regulation & Motivation [*section bias: + convert –ve statement scores using Adjusted Score = 1 – Actual Score*]
2.2-1   I find it quite difficult to concentrate on my work most of the time
2.2-2   I don't think my learning challenges make me any more anxious than anyone else
2.2-3   I use my learning strengths to help me with study strategies
2.2-4   I need to work much harder than my friends to get similar grades
2.2-5   I often feel frustrated when trying to study
2.2-6   I enjoy my studies even more when the work becomes difficult

Academic Self-efficacy   [*section bias: +*]
2.3-1   I believe that my learning strengths really make a difference to my academic progress
2.3-2   I plan and organize my work carefully which I believe helps me to get good grades
2.3-3   I don't think my learning challenges make any difference to the way I tackle my work
2.3-4   I approach my written work with a high expectation of success
2.3-5   I believe my learning strengths help me to be more creative
2.3-6   I can manage my studies quite adequately without any help

Self-esteem   [*section bias: + convert -ve statement scores using Adjusted Score = 1 – Actual Score*]
2.4-1   I often felt pretty stupid at school
2.4-2   If I try hard, I can achieve just as much as anyone else
2.4-3   I think I'm good at studying, perhaps even academically talented sometimes
2.4-4   I approach my written work with enthusiasm
2.4-5   At times, I think that I'm just hopeless at tackling academic work
2.4-6   My contributions in class are usually rubbish, so generally I don't bother

Learned Helplessness   [*section bias: -*]
2.5-1   When I start a new course or topic I usually think it will be too difficult for me
2.5-2   I've had help for dealing with my learning challenges but it hasn't made any difference
2.5-3   I'm generally not surprised when I get a low grade
2.5-4   I will always be held back by my learning challenges
2.5-5   I think my grades are as much to do with luck as with any effort on my part
2.5-6   However hard I try this rarely makes a difference to my grades

Academic Procrastination   [*section bias: - convert +ve statement scores using Adjusted Score = 1 – Actual Score*]
2.6-1   I usually finish my essays or assignments well in time for the deadline
2.6-2   I generally put off getting started on my essays or assignments until I really have to
2.6-3   For one reason or another, I often have to request extra time to complete my work
2.6-4   As soon I'm given an essay or assignment title, I'm usually eager to get going on it straight away
2.6-5   My essays or assignments would probably be better if I didn't have to rush to finish them
2.6-6   I often find other things to do rather than working on my studies

The simulated mean values for each construct for both research groups ND and DI were derived from my own judgment about values that I might expect from the QNRs and these are plotted on the radar plot in contrasting colours, allowing the simulated data from the respondent's QNR reply, also condensed into mean values for each construct, to be overlaid onto the same axes.

As can be seen that in this case from the simulated 3-plot-radar profile below (and here), Respondent ID: 30113372 (which was me, replying honestly), we may conclude

that THIS respondent's profile is more aligned with the mean profile of the non-dyslexic research group than with the dyslexic research group – which is as it should be since I'm pretty certain I don't have a dyslexic learning difference. It is this three-plot overlay that will be created for each respondent from the main research QNR and as outlined above, the comparative analysis of these plots will be forming the basis for attempting to identify students in research DNI, that is, those whose profiles are suggesting a more 'dyslexic' profile than not. Subsequently, Academic Behavioural Confidence will be analysed from the data that each respondent will have completed in reply to the stem statements at the top of the QNR and the complete results will be used to address the main research hypothesis and form the major part of the project discussion. In the profile below, the Academic Behavioural Confidence (ABC) score is included, together with the 'Dyslexia Index' (DI) which I discuss in more detail below. The visualization process for these two constructs is still being explored and so in the final diagrams they may look a little different to the representation here although presenting them as bullet diagrams as shown below is looking promising. These were constructed using a neat snippet of javascript developed by JL Briggs based on the original concept idea from Stephen Few. The base version I modified to create the bullet diagrams in the screenshots below is saved in my jsfiddle.net account here.

I ran a further simulation through the main research QNR, this time trying to put myself in the place of a moderately dyslexic student by gauging how this individual may respond to the stem statements. I replied to the stem-statements based on my experience of working extensively with students with dyslexia in HE. This generated the 3-plot-radar profile Respondent ID 47218304 (below, and here) and to my delight, it seems clear that this time the psuedo-respondent's profile is clearly more aligned with the simulated mean profile for dyslexic students, research group DI.
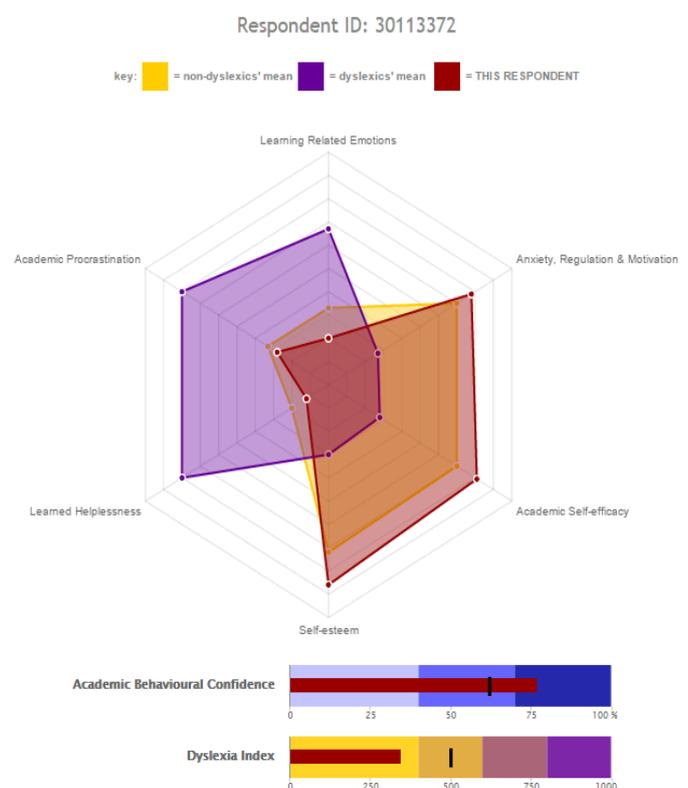
So given these two simulation runs, albeit with QNR data that has been manufactured – for the second one at least – this comparative process that I have designed does now look like it may work, which is highly encouraging and I hope vindicates the careful planning and extensive THINKING that I have put into the project to date.

Creating a 'Dyslexia Index'

Now remember that a fallback position has been built into the main research QNR to cover the possibility that the catalogue of 3-plot radar profiles do not show the distinctive differences that will enable research group DNI to be established:

The final section of the QNR is the last of the Likert scales and contains further Likert response items as stem statements. These have been developed from the preliminary enquiry, DIMENSIONS OF DYSLEXIA, carried out during the Summer of 2015, reported more fully in a separate StudyBlog post and on the project webpages here.

This section of the QNR generates a similar list of return values ranging from 0 to 100 to indicate the level of respondent agreement to stem statements relating to dimensions or attributes of dyslexia that are prevalent in university students. From these, a 'dyslexia index' will be formulated that will be considered in conjunction with the profiles. For the first 'psuedo-respondent' (me, being me) the 'dyslexia index' produced a value of 386 along a range from 0 to 1000 and I do not have a dyslexic learning difference. For the second pseudo-respondent (me, being dyslexic) the 'dyslexia index' is 640. So it is going to be both very important and highly interesting to see the range of indicies that my QNR measure returns for respondents who are known to be dyslexic and in contrast, for declared non-dyslexic respondents, so that this can
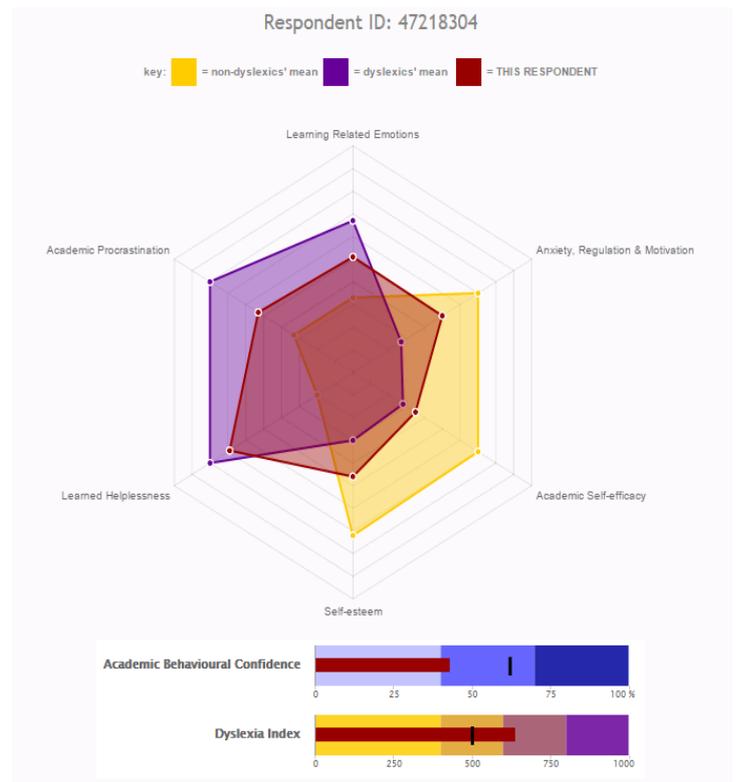
add support to the profiles. Significantly, the analysis of these dyslexia indicies will hopefully enable a tentative boundary value to be assigned for an index point above which a respondent might be declared as having a profile aligned with dyslexia, which, in (my) theory at least, will have been established anyway from a respondent profile that is aligned with the profile of means for dyslexic students.

I will be composing a fresh StudyBlog post that describes how this 'dyslexia index' process has been formulated but reflecting on the information that it should generate has caused me to revisit in my thoughts the thorny question that I alluded to in my original research proposal bid – that is: hOW DYSLEXIC IS DYSLEXIC? This is interesting as current frames of reference used in the HE sector appear to categorize dyslexia assessments into levels of 'severity' of dyslexia, very roughly as discrete categories ranging from 'mild' to 'very severe'. A fresh Literature Review Map will be constructed in due course to summarize research on the psychometrics of dyslexia but a cursory glance appears to show many researchers expressing considerable frustration with a persistent focus on metrics related to deficit.

To complete the post here, it is appropriate to display again and side-by-side the two, 3-plot radar profiles generated by my two psuedo-respondent personae. Even though these profiles are both simulations I did not fiddle the data so that these profiles would emerge so it is encouraging to note that they both appear to tie in with the additional data generated by the QNR: the ABC score, which is going to be the core comparator between each of the three research groups, and the Dyslexia Index, which looks like it may provide an appropriate and highly interesting additional metric.

Also note the random respondent ID numbers which are generated by the QNR form-processing script. This is a very important feature of the data collecting process and are essential for maintaining respondent anonymity whilst at the same time will provide an identifier that will enable a respondent to request revocation of the data submitted if they wish.