

# Analysis for Ron Mitchelson and PPC Committee regarding Department Self-Study Data\*

Report #2.0<sup>†</sup>

Ron,

I've had a first look at the self-study data, the rankings, and your analysis. The errors I find in the self-study data are a problem for the rankings. These data need to be fixed and double-checked for errors. Having done this there are still major hurdles to sensibly compare all units across a single or even a few dimensions. The priority of a unit must naturally depend on the direction and goals of the university. Rather than provide a single ranking that is supposed to apply to any direction that ECU might take, I suggest using the self-study data to provide a clearer understanding of the contributions each unit makes to the university.

Please let me know if you have any questions. Thanks.

Paul

## Contents

<b>1</b>	<b>First Look at the Data</b>	<b>2</b>
<b>2</b>	<b>Rankings</b>	<b>2</b>
2.1	Concerns regarding the rubric - Defining Components . . . . .	2
2.2	An Example - Quality of Instruction . . . . .	2
2.3	Concerns regarding the rubric - Combining Components . . . . .	3
2.4	Relationship to Self-study data . . . . .	3
<b>3</b>	<b>Statistical Analysis</b>	<b>4</b>
<b>4</b>	<b>Recommendation</b>	<b>4</b>

\*Calculations done on data in c:/DATA/ecu/chair/PPC/AllProgramsData/Report on Tue Nov 29 17:20:25 2011 using R version 2.13.0.

<sup>†</sup>Paul Vos [vosp@ecu.edu](mailto:vosp@ecu.edu); 252.744.6040

## 1 First Look at the Data

This has been redacted because, according to emails of 29 Nov 2011 from Ron Mitchelson and members of the PPC committee, the data provided on the website differ from the data used for the rankings. The data submitted by the departments were “cleaned” before being used for the quantitative analysis. Quality checks will be performed on the final data used by the PPC if provided.

## 2 Rankings

The rankings are based on the self-study data and the accuracy of these data can be called into question as illustrated in Section 1. Concerns extend beyond the collection of the data to the definition of the variables themselves and the extent to which these variables measure quantities that allow for comparison of diverse departments in terms of productivity, quality, and centrality. Productivity appears to be one of the easiest quantities to measure. If we consider SCH per FTE we see that there are many factors that are related to this variable: graduate vs undergraduate, class size, number of instructors/grad students providing instruction, clinical vs nonclinical instruction. This is not intended to be an exhaustive list but to indicate the difficulty in using this one variable across different units. Productivity in scholarship can be measured by refereed articles per FTE. However, books also need to be considered and quantity cannot be the only consideration. Should sole authorship be counted the same as a sandwich author among eight other authors? How are musical and theatrical performance factored in? Quality and centrality are even more difficult to measure.

### 2.1 Concerns regarding the rubric - Defining Components

The concerns with the underlying data and with the measurement and definition of the variables is not addressed by using a panel of judges. The judges can be of utmost character and of highest quality but I find it very difficult to see how this rectifies the problems with the data. I think this is a very difficult argument to make. The rubric organizes some of the variables in a meaningful way into 9 components. But there is no indication of how items within a component are combined by the judges.

### 2.2 An Example - Quality of Instruction

The rubric included SOIS and grade distribution as indicators of quality of instruction. Consistently low SOIS for an individual instructor can be indicative of problems but these are of little value to compare departments. The rubric indicates that “Grade Distribution at undergraduate and graduate levels” is used, the self-study data only has mean grade across all sections (graduate and undergraduate). It would be difficult to argue that higher mean grades indicated higher quality of instruction. I would argue that the judges did not have the data necessary to judge the quality of instruction in BIOS nor do they have the expertise. I don’t think BIOS is unique in this regard. Nor is this component unique.

### 2.3 Concerns regarding the rubric - Combining Components

The validity of ratings of other components can be called into question and it must be noted that not all components should carry the same weight. It is difficult to argue that components are equally important, and yet, the rubric treats them as such.

### 2.4 Relationship to Self-study data

Dave Cistola in a 23 Nov 2011 email shows that the size of the department is correlated with PPC Total Score. As you point out, it is plausible that size is correlated with factors that should be related to PPC Score. Listed below is a table of similar bivariate regressions performed by Jason Brinkley. Like the analysis of NFAC, each predictor variable was examined against total PPC score and a transformation was used if the scatterplot showed a distinctly nonlinear relationship. The results are shown in two blocks, the first with variables related to the size of the unit and the second with variables related to the potential quality of the unit. Variables were selected for this table to represent as broad a cross section of departments as possible. We agree with the earlier comments that no single variable explains an extremely high proportion of variation in the total PPC scores, however the R-square values that relate to the size of the department are in general 2-7 times greater than variables that represent the quality of the unit. Especially striking is the idea that the relationship between total SCH and PPC score has an R-square value that is more than double the R-square of PPC score and SCH per Faculty FTE (a measure that controls for the size of the unit).

Variables Related to Size				
Item	Predictor	Definition	Transformation	R-Square
1	NFAC	Tot. No. Faculty FTE	log	0.268
2	FAC\$	Tot. Salary of Faculty	log	0.360
27	SCH	Tot. SCH produced	log	0.355
7	NSPA	No. SPA FTE	log	0.316
12	GAS\$	Tot. Salary State Supp. GA	dichotomized	0.222
29	NUMAJ	Tot. Under. Majors	none	0.065
30	NGrMaj	Tot. Masters Majors	none	0.032

Variables Related to Quality				
Item	Predictor	Definition	Transformation	R-Square
87	MCITEF	Mean Citations/T&TT FTE	square root	0.162
44	MSCHF	SCH/Faculty FTE	none	0.158
90	F&A\$F	F&A/T&TT FTE	square root	0.094
57	MREFF	Ref. Articles/Chaps./T&TT FTE	none	0.055
58	MPRESF	Presentations/T&TT FTE	none	0.044
66	MOSERF	Service Activities/Faculty FTE	none	0.018
76	MSOIS	Mean Unit SOIS scores	PHIL excluded	0.011

Note that GAS\$ was categorized into two groups, those who provide at least some graduate funding and those who do not (either because of lack of funds or no graduate students in the department). It was not always the case that departments with high number of graduate students also provided high graduate funding. Conversely, graduate funding does not mean there were graduate students (FORL data template shows \$15,000 allocated

each year to graduate assistanceships but it also shows that no graduates students were enrolled). Additionally, the relationship between PPC score and mean unit SOIS scores were considered with PHIL excluded because the reported SOIS scores (above 80 each year) could not be compared with other departments.

In your reply to Dave you state “I’m asking each PPC member to take this empirical finding to heart and to help assure that size (by itself) does not influence our recommendations.” The concern is not with intentional bias but the biases beyond one’s control. It is important to recognize the limitations of one’s data and identify questions that can and those that cannot be addressed. The above table illustrates at least some issues that can occur (even if unintentionally) when very different units are scored and then ranked altogether.

### 3 Statistical Analysis

There is great flexibility in the analysis that can be done with data. Often the techniques that are chosen depend on discipline of the researcher. Whether one uses classical methods such as cluster analysis or newer techniques, the conclusions drawn from the analyses will not be valid if there are serious problems with the data. Given the problems described above, any statistical analysis is premature.

### 4 Recommendation

While the desire to have a single ranking, or even three rankings, is understandable, it would appear very difficult to accomplish this given the diversity of the units involved. Furthermore, since the priority of a unit will depend on the strategic direction taken by the university, this raises the question of whether it is even possible to provide a ranking that will adequately address priority for any direction that ECU takes. Even if there is no overall ranking of the 52 units, the data can be used in a productive manner using the following steps:

1. Provide the final self-study data used by the PPC to allow for a thorough vetting of the data.
2. For each of the nine components, PPC evaluates units in terms of whether they appear to be lacking in this component. If so, PPC gives a written justification for this rating based on the self-study data and any other information used by the committee.
3. Units provide a written response to the PPC.

These steps allow for a better understanding by the PPC of what each unit brings to ECU and its mission. Strategic directions that require the elimination or reduction of a unit can be better understood by considering how that unit contributes to each component. This is consistent with the goal of the PPC expressed in the 15 Nov 2011 report:

Our view of program prioritization is a positive one which emphasizes the importance of self-study and reflection that yields deeper understanding of our academic and service programs and the resources needed to support them.