

**Newark Faculty Council (NFC)**  
**Preliminary Response to the Proposal of the Task Force on Evaluating  
Teaching entitled:  
A Proposal to Improve the Evaluation of Teaching at Rutgers University**

We applaud the fact that Rutgers is undergoing a system-wide discussion of teaching practices and evaluation practices for teaching. It is notable that many faculty members are actively engaged in the discussion and eager to be so.

We have found that the taskforce draft proposal does not do enough to address what *good teaching* is. This needs to be articulated. It needs to be addressed both university wide and at the department level. We recommend that each department addresses this issue as part of the self-examination that the task force seems to call for. Each department should formulate a vision of *good teaching* appropriate to its discipline. Below we suggest a template for each department. One example of such a template that has been fleshed out is the Law School's document which is attached below. While creating a document as outlined by the template calls for a lot of work for individual departments, it is a one-time exercise.

**Main Recommendation**

We cite the teaching document from the Law School and we would like to see every department develop a similar document.

- It begins with a statement what the school views as good teaching.
- It identifies the different types of courses the school/department teaches.
- For each of these it lists desired elements.

*We suggest that this could be the Rutgers-Newark standard format.*

The departmental document should be included in every reappointment /promotional pack that is sent out of the department.

Each department should also list (for its own uses only) all metrics it expects to use, what each metric measures and its strength and weaknesses. Faculty should conduct paper mid-term evaluations which are used only for their own edification and not shared with the department.

**Additional Comments**

1. The task force report could go farther in addressing what the various metrics actually measure and concentrate further discussion on the various metrics that are available. There are pros and cons to each metric. It would be a good idea if each department used a

combination of metrics but note the plusses and minuses of each in a written evaluation of faculty. Also evaluation of a faculty member's teaching should be given as a narrative and not numerically.

2. The NFC discussions raised specific criticisms of the SIRS survey. If Rutgers continues to use it (use it as one factor in evaluating teaching), it can and should be amended and its use put into a proper perspective. Procedures must be put in place to assure vast improvement in the student response rate. We note that some studies indicate that student surveys do not measure what is learned. Some of the NFC discussion comments relevant to SIRS are also attached below. But the NFC is also concerned that these comments are not used as base for *cherry-picking*.
3. We would NOT like to see numerical rating attached to the *list of desired elements*. We recommend a deeper and more thoughtful evaluation.
4. We note every evaluation tool has strengths and weaknesses and even a set of tools used in combination will have flaws.
5. There are questions about how the system will address the problem of teachers who are found to be weak. Weak teachers can result in the department's instructional responsibilities being carried unequally. This can result in other faculty in the department having reduced time and energy to devote to research.
6. There is concern about the training for teacher evaluators, their potential bias, and also about the time consuming aspect that can be involved for evaluators. Evaluations should be done by senior tenured members of each department.
7. Statements from various discussions including those that are quoted in the Oct. 16<sup>th</sup> minutes are included below with the proviso that we don't want changes to SIRS that are made to address our concerns cherry-picked.

Some of the points are summarized here.

"praised the proposal but noted that we need more than "two questions" on the SIRS student surveys. Questions on level of course difficulty and work load must also be included. Also, the two questions recommended are "yes/no" questions. This is silly. At the very least, the answers should be on a scale. Finally, the language in the TEC report blames faculty for the low participation in SIRS surveys. Obviously the problem is the movement online of the surveys. ... Proposed that we find a way to do the surveys in class, during class time, through wireless connections and also asked about teaching observations. Who observes who? Would PTLs observe full professors? What would the policy be on that? Would we have a policy for remediation if the teaching report is negative? How would we help teachers improve?"

"noted that it was foolish to make the student comments voluntary. We need their comments for faculty development"

"we adopt a question: "How would you rate your own contribution to the work of this course?", ... found the student responses to this question to be honest. ... also has a written mid-semester response on how the course is going, which we could all adopt. "

“Suggest: we look at how other universities rate teaching and there are some documents on this that could be shared with us.”

“suggest a standard subset of questions for the student evaluations to which other questions can be added. We must have student answers be on a scale (not yes/no). We should be able to compare across departments that way. The standard questions should be balanced, but also tailored to specific disciplines (labs, seminars, etc.). It was noted that that in (CMBN), only the director could undertake the teaching observations. .. this would produce an enormous workload for the director. There could also be a problem (if the director doesn't do all evaluations) of peer evaluations in which the evaluator is not himself/ herself a skilled teacher. Would the person evaluated have a chance to respond? Can we accept that teaching portfolios will vary by department?”

“ flaws in the SIRS surveys (if we did medical studies as we do SIRS surveys, people would die). .. blamed SIRS for a decline in teaching quality, as people teach easier courses in order to get better responses on SIRS. We need higher response rates. We should copy other universities' methods, for example by delaying final grades until the SIRS survey is in. ... also suggested a breakdown for each course on how the grades are given (percentage of As, Bs, etc.). When outsiders evaluate teaching, they look at graded papers, etc – not just at the syllabus. Will our evaluators be willing to do this? He once took over two sections of 150 students, and found that they had been evaluated incorrectly.... noted that a star teacher would do well in research as well.”

“ ...in SIRS evaluations, ratings tended to be lower for mandatory courses, for courses graded on a curve, and for very large classes ... faculty has not been involved in writing evaluations or giving teaching feedback... suggested that faculty across the university confer on how to give proper teaching feedback.”

“ if we did a large document about teaching, its suggestions could be cherry picked by the Ad Hoc Committee or the administration . How do we evaluate without SIRS?”

“we need more than SIRS for evaluations”.

“we should state that SIRS only measures a percentage of teaching strengths. We need additional measures.”

“ warned that we should be comparing types of classes (rather than classes in general). It's easier to get good evaluations in smaller and more exciting courses.”

“evaluations by peers can be unreliable. Faculty members are not trained in what to look for. You'd need two evaluators for each one. We don't have a consensus on what makes a good teacher. For example, some see lecturing as inherently bad, while others are masters at lecturing and should not be graded down because they lecture. We shouldn't privilege one form of pedagogy over another.”

“in some fields, only the faculty director would have the expertise to do the teaching evaluations”

“one year, as undergraduate director, did about a dozen teaching evaluations of PTLs and TAs (this hadn’t been done in several years, so the faculty member had to do all of them). On one hand, doing so many made it possible to compare who was and was not an effective teacher. On the other hand, it was hugely labor intensive. No one should be asked to do this sort of work without either financial payment or a teaching reduction.”

“ the TEC report mentioned in passing that SIRS and other standardized evaluations often express unconscious student biases against women teachers and teachers of color. What will be done to correct for those biases?”

“agreed that doing many teaching observations was a lot of work.”

”teaching was reviewed each semester by a different person. We should articulate what makes a good teacher.”

“all methods have pros and cons. We should each articulate what makes good teaching. .. do people with weak evaluations ever improve as teachers.”

“ yes, they improve. If we didn’t believe in that we would not be teachers.”

“some SIRS questionnaires have a deadline before exams, and some after. This can distort results”.

“An event in previous system in place in one department of teams of observers (one junior, one senior) summarized. In this case, the senior evaluator remarked afterward to the junior evaluator that it was worse class ever observed. Yet when the junior faculty was asked to sign the evaluation written by the senior faculty member, it was a glowing report. Reports of class observations now tend to focus on basic teaching capacities questions (is the voice level appropriate? Is the board writing legible?) . “

“Evaluators needed to be trained to detect their own biases before evaluating others. “

“Departments themselves should develop guidelines for assessing atypical courses, such as the art/life studies courses mentioned at our meeting, group-taught courses, “boot camps" as well as tightly focused and interactive graduate seminars. These are examples of "non-typical" teaching that may differ widely from what would be considered necessary components of the typical course curriculum and vary tremendously by department.”

“there is serious concern about student comments not remaining an integral part of what is communicated back to the faculty”

“One should make sure teacher evaluators are trained and that this is done in pairs with credit for performing this service”.

\*In view of the short time-frame given for a response, this response is a preliminary response. It may be up-dated during the Spring 2018 semester. We are concerned that the very short deadline for this important University-wide task will not lead to a thoughtful and good outcome.