

# Doing the PIF

## A Critical Review

### Summary

YUFA members have indicated through the survey they participated in earlier this summer that **the AAPR has been a massive failure**. They provided an overwhelmingly negative assessment of the Program Information Forms (PIFs) and the process of completing them. They complained about the PIF methodology and, often bitterly, about the quality of data supplied by the administration that they were expected to comment on. They found the additional workload to complete these tasks appallingly heavy. They have little or no confidence that the AAPR will deal with their programs fairly. This exercise has undermined faculty morale at York University to an unprecedented extent. There is no reason to have any faith in any reports based on this deeply flawed process.

### A Brief History

In May 2013 the Provost and Vice-President Finance and Administration announced a process of program prioritization. It was soon clear that the administration had sought out the expertise of the US management consultant Robert Dickeson in designing the proposed survey of York programs. In October the President formally announced a renamed Academic and Administrative Program Review. The launch of this initiative was delayed as dissent emerged in Senate. The process was then altered to allow for the election of two Senators to the sub-committee drafting the template to be used across all academic programs and for some time in Senate to comment on the template, by that time known as the Program Information Forms (PIFs).

These forms were distributed at the end of March, and had to be completed by the end of June. Across the campus, “authors” were assigned to complete the forms, and in some cases, where an individual unit had many programs, they sub-contracted the responsibility to “collaborators” for specific programs. Statistic data were provided for each PIF by the central administration, and authors and collaborators were expected to comment on these “metrics.” Training sessions were held, and numerous academic administrators in deans’ offices became consultants to those completing the forms. All PIFs were reviewed in decanal office before final submission. One faculty, we learned, hired a professional writer to prepare its PIFs.

The Academic PIF had eight topics that asked for short answers to large, wide-ranging questions. For each topic there were between two and seven individual questions, making a total of 35 questions, some of which had two or more parts. Each topic was awarded either 10 or 15 points, and there was a limit of 300, 400 or 500 words on all the answers for each topic. In total, the maximum length was 3,500 words – so the average available per question was 100 words, though this varied from one topic to the next.

At the end of June, two task forces began the work of examining the PIFs and assigning them numerical grades. The Academic Task Force has sixteen members selected by the president. The three faculty members from LA&PS on that body dramatically under-represent the Faculty that accounts for 104 of the 257 programs being assessed, while the two faculty members from Law over-represent a faculty with just five programs being assessed. The overall project Steering Committee includes the Deans of Business, Health and Environmental Studies, but no representative of the Faculty with the largest number of programs by a wide margin.

## **A Methodological Critique**

In late March, YUFA approached the leading expert at York on survey research, Michael Ornstein, former director of the Institute for Social Research, and asked him to assess the Program Information Form designed for academic programs. He made a verbal presentation to a YUFA forum on prioritization and subsequently prepared a written version (“*The Academic Program Information Form for York University’s Academic and Administrative Program Review*”).

He began by noting the limited space provided for answering dense, complex questions, and wondered how authors would respond to large questions in the small space provided. “While it is possible to give *some* answer ... in 150 words, the question is so complex that it requires choosing among many, many things one could say. Clever respondents would think ‘What does the *Task Force* want to hear?’ or ‘How can I make the best case for my program’ and write that; and good writers would be eloquent. The clever and eloquent answers, however, would each address different issues in different ways. The question is so broad that each program and each individual ‘author’ approaching the question would effectively be answering a different question. This strikes at the heart of the idea of ‘standardization,’ which is *the* foundation of valid and reliable survey measurement.... Ambiguity and complexity are the enemies of good survey measurement.” He noted that quite different answers could be prepared for a particular question, and that there was no obvious and fair way to determine which one was more worthy of a high score. He went on to question how many programs could supply the kind of information on “output” that was being requested, other than the large professional programs that tracked their graduates.

He concluded by asking: “Is it possible to complete the entire *APIF* questionnaire? The answer is ‘yes,’ of course. Indeed, similar exercises have been completed at many universities and there is a mini-industry devoted to conducting these surveys. The fundamental problem is that high rankings on the *APIF* survey will have as much to do with the cleverness of the questionnaire ‘authors,’ the pre-existing structural advantages of some programs, and the interpretation of the results by the Task Force as the true merits of programs.” He stated “unequivocally that the questionnaire is no basis for credible research.” He also worried that such a shoddy research tool would “have real consequences for our programs at York.”

## **LAPS Faculty Council Weighs In**

On 8 May 2014, the Faculty Council of Liberal Arts and Professional Studies passed the following motion:

That the Faculty express the grave concerns and serious doubts of its members that the concept, method, data collection and analysis associated with this review will provide either reliable or valid information upon which sound academic/administrative decisions concerning the future of the Faculty can be made.

## **Surveying the Authors**

Between April and June, the YUFA office received many calls of concerns from members who were caught up in the PIF preparation project. The YUFA Stewards Council had struck a Working Group on the AAPR, and, early in June it decided to circulate a survey questionnaire to all YUFA members who were listed on the AAPR website as “authors.” We also asked that those people provide us with the names of the “collaborators” they had worked with. In most cases, that meant the people who took responsibility for a single program (e.g., GPDs who worked on MA and PhD programs). Sometimes, authors supplied us with names of colleagues who helped draft the PIF answers (as appeared to be the pattern in several units). In all, we reached about 220 full-time faculty members. We heard back from 137. This is what they told us:

- 88% thought there had been insufficient consultation with units in developing the AAPR.
- 65% thought the time set aside for completing the PIFs was inadequate.
- 63% lacked confidence that their programs will be treated fairly in the process.
- 90% were worried that their programs will lose resources.
- on average, respondents each spent more than 44 hours in the process of completing PIFs.

In July we did a follow-up survey that asked further questions about the quality of the data that had been supplied and the methodology used. We received 40 responses.

- 90% thought there were problems with the statistical data provided by the administration with the PIFs (45% serious problems, 40% minor problems).
- 80% used the data only partially and supplemented it with additional data; only 10% used it fully, and 7.5% did not use it at all.
- 85% thought there were problems with the methodology (75% major problems, 10% minor problems).

## **Predominant Issues**

The two YUFA surveys asked some open-ended questions. In the first, a technological glitch in the software led to the truncating of lengthier answers, but this was not a problem in the second. Virtually all respondents are included in the following extracts from the written comments. With only a few exceptions, each comment came from a different respondent.

### ***Positive Responses***

A small percentage of those surveyed had at least some confidence in the AAPR process. Some felt their programs were strong enough to weather the ordeal, and they expected to be winners in the process. Others expressed confidence in those who have been mandated to undertake the prioritization. These voices were, however, drowned out by the more negative chorus.

“The financial health of the program is quite good and benefits the Faculty as a whole.”

“The program has a strong, professionally based curriculum and is in demand.”

“Our program performs well on most of the parameters being considered.”

“I believe we have a strong program that meets the criteria for sustainability....”

“The program I represent helps York University secure its leadership and reputation in the field.”

“I know some of the people on the committees and I respect their integrity and judgement.”

“The colleagues that will be reading these will try to be fair.”

“I have trust in my colleagues on the Task Force.”

“I believe that the colleagues who will be reviewing the documents will do their job.”

“The evaluation process is fair and relatively objective.”

“I trust that the evaluation process will be conducted fairly.”

“I think that programs will be assessed on their merits and failings and decisions will be made for the overall common good.”

“I think we’ll be treated fairly because we poured a lot of resources into preparing the best case.” [this comment came from the faculty where a professional writer was hired to prepare the PIFs]

### ***Quality of Data***

There were many complaints that the “metrics” supplied by the senior administration were inaccurate or incomplete, and that some arrived late in the process (in the case of some FGS

data, revised information was sent out the day before the final deadline for PIFs). The exercise revealed the chronically inadequate data collection that the university undertakes and the difficulty of making valid decisions based on what is available.

“New and ‘updated’ data continued to be sent to authors up to the last minute, and for some, after the PIFs had been submitted for approval. There were many gaps and errors in data – even basic data on student enrolments. Program enrolment data we received from the Faculty for the AAPR were different (often lower) from data generated by the same source only a few months earlier.... When inquiring about how to handle the data gaps and discrepancies, we were advised by one source not to worry about the numbers so much – that we should instead focus on the narrative aspects of the survey questions. A different source advised that program numbers would be the single most important factor in the evaluation of program efficiency and sustainability. If the latter is true, it is most disconcerting that complete and reliable data could not be provided by the University. Even more disconcerting is that all of the incomplete and unreliable data will in turn be used ... to justify administrative decisions about the value and validity of academic programs.”

“The abundance of data was, in itself, somewhat overwhelming, especially given the tight timelines for the review. Discovering and gisting what was relevant in the barrage of numbers was often very challenging.... Anomalies or apparent small errors in the numbers were distracting and ultimately a waste of time to try to correct.... Data continued to arrive right up until the submission deadline for the PIFs. It was obviously impossible to redraft PIFs in light of late-arriving data which required time to digest, etc.”

“I’d like to know the source of some of it, as some was wrong – in the case of the financial data, the Dean’s office was clear that they could not divide costs according to program to generate real, accurate data, so they used a predetermined formula that someone made up and that many could not understand. This fact kind of negates the entire value of those data and their interpretation I should think.”

“The aggregate data was suspect and we spent a great deal of time reproducing it by hand to check. The research data from the Faculty had serious errors (missing professors, grants, etc.) and the Faculty total column was off by a factor of 2 for the grant money and 3 for the publications.”

“As ORUs, we received very little information from the employer, and what we did receive was quite arbitrarily assembled. In particular, we received no information at all for the faculty and staff complement form, nor any guidance for how to complete it, so we all did completely different things, making the information impossible to compare across units.”

“We were deluged with data, some of it incorrect, none of it properly contextualized.”

“The financial data are inaccurate.”

“The input data is horribly incomplete ... [and] extremely poor quality”

“Data collected have serious flaws.”

“Data provided is not correct and my efforts to seek corrections have been futile.”

“The data used [was] often outright incorrect.”

“The information supplied to us was incorrect, beginning with our faculty complement.”

“We received data that in its majority seem to have been incorrect.”

“The information generated doesn’t take into account the hybrid structure of many York programs.”

“Our discipline can turn to none of the external metrics that other programs can.”

“Data are not sufficient to complete the PIF.”

“Data was unreliable and incomplete.”

“The metrics are flawed.”

“The data ... was full of some serious errors. I had to spend many extra hours to make sure that the data we input into PIF corresponded with the reality.”

“Data was inaccurate at times and certainly open to serious questions in other respects.”

“It was not up to date.”

“The way statistics were presented made it virtually impossible to form a clear description of the value [of programs].”

“The data sources provided were poor.”

“The form asked for information that either made no sense or was not provided (or both).”

“The whole process revealed that the university does not have correct data on many fronts.”

“Data provided was completely inaccurate (included data from another unit).”

“It was extremely difficult to obtain accurate information about programs.”

“Data [was] not accurate or not all available; incomparable between programs.”

“It was frustrating to keep getting new data as the process went on.”

“There were changes at various stages to the dataset which caused some confusion.”

“The data required was only provided a week before the deadline.”

“The university does not keep the stats they were asking us for and then made us use only the minimal data they had.”

“Data is often aggregated for larger units, and not pertinent to component units.”

“If they had provided some definitions of the terms and when they collected the data, it would have been much more helpful to read and understand them.”

“Analysis based on discrepant (and in some cases clearly incorrect) data is meaningless.”

### ***Methodology***

Many respondents expressed misgivings about the overall methodology of the review process. The definition of a “program” was problematic. The tight framework of common questions for everyone frustrated many authors. The relationship between the centrally distributed “metrics” and the brief qualitative answers was also frustrating for many. The word “flawed” appeared repeatedly.

“This is perhaps one of the most methodologically unsound surveys we in the wider social sciences (including the more flexible interdisciplinary social sciences) have ever seen. It is especially embarrassing that it has been produced and endorsed by the University. This is in large part due to the extreme scope of the survey and the attempt to create standard questions that could be meaningful across and within disciplinary and interdisciplinary fields. In the attempt to capture everything and appease the many concerns raised about the methodology along the way, the survey is rendered meaningless.”

“I did not have the sense that the forms were developed with the research expertise that York faculty [could have provided]....”

“Too little collegial consultation at the beginning and too little training.”

“Lack of consultation about relevance of much of the material.”

“The purpose, theoretical grounding, and method were opaque.”

“The process was flawed.”

“Flawed methodology; contradictory and redundant questions; questionable rubrics; suspect assumptions....”

“The validity and reliability of the exercise is deeply flawed.”

“The methodology is flawed, ambiguous.”

“Methodological madness – a hoax.”

“I don’t think the methodology is transparent. Efficiency does not mean quality or good practice.”

“A very broad paradigm is being applied equally to all programs, whether they be applied sciences or [other disciplines].”

“The time allotted for the exercise was problematic. The definition of programs was perhaps too narrow and there seemed to be little appreciation of the nature and type of programs and their unique qualities.

“Probably not sensitive enough to interdependence of many grad programs’ faculty and curricula.”

“Program is not reviewed in all aspects”

“I have no faith in the underlying logic and methodology of this process whatsoever.”

“Training was provided but very badly.”

“The time allocated to the process was not adequate.”

“It is not an academic evaluation of programs by peers.”

“I was given only a few days to incorporate feedback on the first draft.”

“It appears that the AAPR was driven by a misguided notion that programs could be evaluated purely on a quantitative basis.”

“The AAPR process seeks to assess the quality of the faculty members. In this process, one of the most important indicators of their quality is taken to be whether they have received research grants. An implicit view is that: the more the number of grants and the more the amount of money one has received, the better is his/her scholarship. Why should this be the case? Is there really a necessary relation between scholarship and getting money?”

“The amount of time and money invested in the ‘support infrastructure’ for this exercise could have been [better spent]....”

“[My] main concern is the comparability of the distinct programs.”

“There was not mention in the methodology regarding programs that are fully linked, such as an undergraduate program that relies heavily on the graduate program.”

“The AAPR does not require or make adequate room for a discussion of the key fact that programs exist within the logic of departments, schools, and faculties. Often, when

taken in isolation, a program is inexplicable or its ostensible lack of sustainability masks its integration into a network of mutual interdependence that cannot be excised.”

One of the central concerns expressed about the methodology of the AAPR was the nature of the questions asked. Respondents found them ambiguous and repetitive, noted that they often didn’t apply to their programs, and thought they gave far too little space to adequately present the strengths of programs.

“The questions were ambiguous at best (too many sub-questions); that made answering in the word limit difficult. I always felt I was leaving something out, or, at other times, I felt some answers had to be repeated in several questions.”

“Obviously, too little space to answer, and too many questions hidden within one question. The questions didn’t allow for qualitative issues to be well addressed.”

“Short answers to complex questions cannot possibly describe my program’s strength and long-term direction....”

“The questions are not designed to assess academic quality, but efficiency from a financial perspective....”

“Some questions were poorly expressed. Too much jargon, too much meaningless verbiage....”

“There are too many questions.”

“Too many questions, many of which brought forth the same responses.”

“The questions were often irrelevant.”

“I don’t think there was enough room to expand upon questions in a comprehensive manner.”

“There were questions which didn’t apply to my program which were unanswerable.”

“The frame of the questions is not really adequate to capture the specificity of my program.”

“The questions were often totally inappropriate for some or all of the units for which I was filling out {PIFs}.

“The quality of PIF depends much on the ability of the author to present information in the best light.”

“Questions were not coordinated with each other. They needed better focus and simplicity.”

“Evidence of success specified in areas such as student quality had no relevance to our program.”

“The explanation of the question is much longer than the space given to answer them.”

“There is general concern that the units were not able to express the strengths of their programs given the restricted nature of the questions, that the tool was not appropriate for a very diverse pool of programs, and that as a result the programs will suffer.”

“Often the questions were three times as long as the allowed answers and it was impossible to explain or contextualize our answers. I am not sure that the answers we were able to provide truly captures what and how we do what we do in our department.

### ***Competitiveness***

Several respondents worried about the competition between programs that the AAPR set up, to the possible detriment of smaller programs in particular.

“It felt as if we were competing against other departments and programs at York and that was not a healthy spirit to encourage.”

“AAPR seems to be an exercise in trying to sell your programs.”

“The future of small programs may be at risk.”

“It is unclear to me what the comparators are for my program.”

“The entire AAPR process has felt like a punishment and fault-finding exercise of the worst sort.”

“The methodology is flawed and does not accurately allow for differences between disciplines and departments.”

“I fundamentally disagree with the competitive nature of this process.”

“What would be the future of Liberal Arts and of the small units?”

“The process seems prejudiced against smaller units whose functions are not easily comparable to others.”

“The rankings for small units, when comparing to large units, is like comparing apples and oranges.”

### ***Evaluation***

Many respondents worried about how the PIFs would be evaluated.

“A clear lack of information about the process of evaluation.”

“Lack of clarity about how results would be used.”

“[I am] not sure about the criteria put in place to evaluate programs.”

[I am] not sure that the evaluation team is representative.”

“I am not confident the reviewers will catch the nuances of issues specific to our program.”

“[I am] unclear on criteria on which judgments [will be] made.”

“[The] function of weighting [was] not clear.”

“[I am] not sure that any outcome is attached to any particular rating in this process.”

“[I] can’t imagine how the members of the reviewing committee will agree on acceptable standards.”

“It would have been useful to know what the exact standards are for the AAPR exercise.”

“I find it hard to believe that the review committees will have time to read and digest all the PIFs.”

“The most basic questions, like how units will be compared to one another across significant structural differences, have not been answered.”

### ***A fait accompli***

Some respondents expressed a belief that the AAPR exercise was pointless since senior administrators had already decided the fate of their programs.

“It’s our program’s feeling that a decision about our program has already been made....”

“It is clear that decisions about programs I am involved in have already been made ahead of the AAPR.”

“My feeling is that the PIFs will be used selectively to support predetermined administrative and academic [decisions]....”

“Our dean’s office has been quite clear about priorities going forward.”

“There was ... a strong temptation to suppose that the point of the exercise was to deflect blame on the PIFs for decisions to be made (possible already pencilled in) by senior administration.”

“From a number of sources, it has been reported that decision about which programs will be strategically promoted and supported and which are considered too small or

costly and at risk of being closed or merged, are already known, and the evidence gathered from the AAPR would be interpreted so as to justify the desired outcomes.”

### **Workload**

Most authors thought that completing PIFs was an unfairly demanding responsibility at a time of year when grading intensifies and research activity, including conferences, normally accelerates.

“This exercise was not figured into the workplan in 2013....”

“The workload generated by this exercise in no way justifies benefits (if any) derived.”

“The time involved was enormous.”

“The time and energy spent will have no impact.”

“The exercise was very time consuming not only for myself, but also for the centre coordinator.”

“The amount of time dedicated to this ‘process’ is a research killer.”

“This exercise took me away from my research activities during a prime season for research.”

“I was preparing [the PIF] at the same time as I was working on some major research proposals.”

“Such a huge commitment from Faculty and Staff above and beyond what is already expected of us.”

“The timing of the deadlines was difficult, coming at the end of the academic year....”

“I spent 4-6 hours a day on my family ‘vacation’ the week before the PIF was due (June 21) on AAPR. This is in addition to the 2-3 months prior on the process and the 12 hour day I spend on the 21<sup>st</sup>. “

### **Morale**

Many respondents found the whole process demoralizing.

“This process wasn’t designed to treat any program ‘fairly.’ It wasn’t about fairness, it is a cost-cutting exercise that is attuned to the cost of everything and has no interest in the value of anything beyond the financial bottom line (e.g., intellectual, equity).”

“The manner in which it has been designed and executed has only fostered or validated feelings of distrust and lack of confidence in the administration.”

“[I am] suspicious of the motivation for this exercise.”

"[I] just wish to reiterate what a waste of time and resources this exercise has been."

"The process has demoralized faculty, many of who feel disrespected...."

"This seems to be a very top down exercise whose purpose is to force more 'efficiencies' on [units]."

"The whole process is an enormous waste of time meant to give justification for the upper level [administration agenda]."

"This exercise was thrust upon us without asking us if it would be helpful in balancing the University budget."

"It was a massive waste of people's time."

"This sets a dangerous precedent about university governance."

"The AAPR survey could be used for unpredictable purposes."

"The only paradigm behind the AAPR is money."

"I am suspicious this is ultimately a budget exercise and money talks. Reducing higher education to a balance sheet will short change the contribution higher education makes to society. It will also curtail creativity in faculty and the population."

"I have doubts about the objectives of the whole process and concerns about the consequences."

"There has been a long history of YU of administrative half truths, unfair policies, and remarkably curious decisions with respect to our grad and UG programs (cannot speak for others). Hard to see past a tradition that has created so much negative ponderings, unproductive thinking, doubt, and lack of confidence."