Report
The Collection and Use of Research Metrics for Assessment and Evaluation

Joint Committee on Metrics
January 2018
Report
The Collection and Use of Research Metrics for Assessment and Evaluation

Prepared by the Joint Committee on Metrics,
University of Manitoba and University of Manitoba Faculty Association
January 2018

Joint Committee on Metrics Members:
Charlotte Enns (Co-Chair), Sherri Vokey (Co-Chair),
Fletcher Baragar, James Blatz, Colin Garroway, Peter Nickerson
Table of Contents

Background: Joint Committee on Metrics.............................................................................................................................................. 1
Summary of Committee Activities......................................................................................................................................................... 2
Principles of Agreement........................................................................................................................................................................ 4
Recommendations .................................................................................................................................................................................. 8
Selected Bibliography on Bibliometrics........................................................................................................................................... 9
Appendix. Appendix H........................................................................................................................................................................... 19
Background: Joint Committee on Metrics

The Joint Committee on Metrics was struck at the request of the University of Manitoba and the University of Manitoba Faculty Association (UMFA) during collective bargaining in 2016. The purpose of the committee was to decide whether the language proposed in the Letter of Understanding (known as Appendix H and found at the end of this report) should be included in the Collective Agreement.

The University and Faculty Association were to each appoint three members to the Committee. The committee was comprised of the following members:

<table>
<thead>
<tr>
<th>University-Appointed Members</th>
<th>UMFA-Appointed Members</th>
</tr>
</thead>
<tbody>
<tr>
<td>James Blatz, PhD, PEng</td>
<td>Fletcher Baragar, PhD</td>
</tr>
<tr>
<td>Faculty of Engineering</td>
<td>Faculty of Arts</td>
</tr>
<tr>
<td>Professor, Civil Engineering</td>
<td>Associate Head &amp; Associate Professor, Economics</td>
</tr>
<tr>
<td>Charlotte Enns (Co-Chair), PhD (EAF&amp;P)</td>
<td>Colin Garroway, PhD</td>
</tr>
<tr>
<td>Faculty of Education</td>
<td>Faculty of Science</td>
</tr>
<tr>
<td>Associate Dean (Graduate &amp; Research)</td>
<td>Assistant Professor, Biological Science</td>
</tr>
<tr>
<td>Peter Nickerson, MD, FRCPC, FCAHS</td>
<td>Sherri Vokey (Co-Chair), MA, MLIS</td>
</tr>
<tr>
<td>Rady Faculty of Health Sciences</td>
<td>Health Sciences Libraries</td>
</tr>
<tr>
<td>Vice Dean - Research</td>
<td>Head &amp; Associate Librarian, Neil John</td>
</tr>
<tr>
<td></td>
<td>Maclean Health Sciences Library</td>
</tr>
</tbody>
</table>

1 University of Manitoba & University of Manitoba Faculty Association. Appendix H: Letter of Understanding Re: Joint Committee on Metrics. Available from: http://www.umfa.ca/images/pdfs/member-resources/Appendix_H.pdf
Summary of Committee Activities

Committee Meetings
The committee met six times between April and December 2017, on the following dates:

- April 3
- April 27
- May 26
- July 17
- October 26
- December 6

Committee Activities

Summer Student
A graduate student from the Archival Studies M.A. program, Christopher Kshyk, was hired from May to August 2018 to complete a comprehensive literature review on the use of metrics to measure the quality of research and scholarly work.

SSHRC Webinar
Several committee members participated in the SSHRC webinar on October 26, 2018 titled “Assessing Impacts in the Humanities and Social Sciences”, led by Peter Severinson and David Phillips.

Invited Speaker
A public presentation by Dr. Vincent Larivièreme, CRC on the Transformations of Scholarly Communication, was organized for all University of Manitoba faculty members on November 23, 2017.

Bibliography
The bibliography was initiated by the summer student and completed by co-chair, Sherri Vokey, to provide a detailed database of publications related to the evaluation of research metrics.

Committee Vote & Summary
The final committee meeting on December 6, 2017 included a vote on the following motion:

“To move to recommend the immediate addition of all language included within section 7 of Appendix H (19.D.1.8.5. through 35.5.7) of the 2016-2017 University of Manitoba – University of Manitoba Faculty Association Collective Agreement.”
The motion failed with three members voting for, and three members voting against the inclusion of the proposed language in Appendix H.2

<table>
<thead>
<tr>
<th>NO</th>
<th>YES</th>
</tr>
</thead>
<tbody>
<tr>
<td>James Blatz</td>
<td>Fletcher Baragar</td>
</tr>
<tr>
<td>Charlotte Enns</td>
<td>Colin Garroway</td>
</tr>
<tr>
<td>Peter Nickerson</td>
<td>Sherri Vokey</td>
</tr>
</tbody>
</table>

Note: The committee felt it was important to state that the vote result did not reflect the unanimous agreement amongst all committee members regarding the issues of bibliometrics, and that the disagreement centred around the specific language being proposed for the Collective Agreement. The attached Principles, Recommendations and Bibliography can provide guidelines for future work in this area.

---

2 University of Manitoba & University of Manitoba Faculty Association. Appendix H: Letter of Understanding Re: Joint Committee on Metrics. Available from: http://www.umfa.ca/images/pdfs/member-resources/Appendix_H.pdf
Principles of Agreement

After careful consideration of the scholarly literature and consultation with experts in metrics and research assessment, the Joint Committee on Metrics is unanimous in its agreement with the following ten principles.

1. **Regular assessment and evaluation is important for both guidance and career progression.**

The committee agreed that proper assessment and evaluation of individuals for tenure, promotion, and performance evaluations is critical for the success of faculty, librarians, instructors, and the university. It was agreed that peer review is the foundation of assessment. The committee further agreed that empirically tested metrics published in the peer-reviewed bibliometric literature that were demonstrably objective and unbiased measures of well-defined aspects of research quality could serve as an additional important component of evaluation. The committee discussed hypothetical scenarios where individual biases of those performing peer review may have negative effects on tenure and promotion evaluations. There was some disagreement about whether the peer review system was inherently biased or whether it was individuals within the system that biased the process of peer review. No clear pathway through which metrics could be used to correct individual biases was identified.

2. **Bibliometrics should be objective.**

The committee agreed very early after formation that journal level metrics (i.e. JIF: Journal Impact Factor) cannot indicate the quality of individual articles in a journal. The committee agreed that there must be an empirically demonstrable and clear relationship between a metric and a precisely defined aspect of the quality of an individual's research for it to be usable for tenure, promotion, and performance evaluations. This relationship between metric and quality should be supported by up-to-date research in the peer-reviewed bibliometric literature.

For example: metric X is being used because it accurately measures research quality Y as demonstrated in the peer-reviewed bibliometric research literature Z. Custom-designed metrics are never to be used until thoroughly tested and assessed. The San Francisco Declaration on Research Assessment’s (DORA)\(^3\) recommendation on this issue is as follows: “Do not use journal-based metrics, such as Journal Impact Factors, as a surrogate measure of the quality of individual research articles, to assess an individual scientist’s contributions, or in hiring, promotion, or funding decisions.”\(^4\)

3. **Bibliometrics should be discipline-appropriate.**

The Committee agreed that there is no logical connection between the volume of publications produced and their quality or importance. It was agreed that after scholarly works are read and assessed via the peer-review process, that it may be desirable or sensible to assess the volume of output respective to those publications (or any other unit of output) if they are deemed to be

---


of approximately equal quality. This is a customary part of the peer-review process that would remain unaffected by the addition of the proposed language found in Appendix H (see article 19.D.1.8.4 on page 63 of the 2016-2017 Collective Agreement)\(^5\). This is does not contravene the language included in 19.D.1.8.1 on page 63 of the 2016-2017 Collective Agreement: “…quantitative analysis and statistics used to assess the quality of a body of work”. The additional language in Appendix H that is being voted on for addition into the Collective Agreement is thus not related to this common task.

4. **Bibliometrics should not be biased.**

The number of citations a paper receives is nearly entirely predicted by the size of the sub-field within which it is published\(^6\). Therefore, used in this way, citations are not an indicator of quality of an individual piece of research’s quality.

A researcher’s cumulative citations are best predicted by the length of time they have been doing research and the number of papers they have published. Time spent doing research is not an indicator of the quality of an individual’s research.

Citation rates and the citation life (time to integration and recognition) of papers vary widely by discipline and sub-discipline. Citation peaks vary across sub-disciplines, and most often many years after publication. Thus, the timing of tenure and promotion applications will happen too early for papers in the majority of disciplines to be recognized and cited. This metric thus negatively affects researchers in fields where citations accrue more slowly, such as in the Arts and Humanities.

The Committee agreed that the citation explicitly named in the language of the Collective Agreement (i.e. 19.D.1.8.1) is a flawed metric that should not be used. Further, the Committee noted that it was unfortunate that the \(h\)-index was explicitly named as an example of a metric that could be used in the language of the Collective Agreement and that this was not eligible for discussion, revision, and vote.

5. **The \(h\)-index and its relatives do not indicate the quality of an individual researcher.**

The committee reached early agreement about the issues and defects associated with the \(h\)-index. Gingras described the \(h\)-index as being like a bad thermometer that can’t go down and can only go up very slowly.\(^7\) The \(h\)-index is highly correlated with years in a position and limited by the number of papers published. It also suffers from all biases associated with citation and paper counting as these are the components of the metric. The Committee noted that it was unfortunate that the \(h\)-index was explicitly named in the language of the Collective Agreement (i.e. 19.D.1.8.1) and that this was not eligible for discussion, revision, and vote.


6. **Bibliometrics alter the focus of research in undesirable ways due to incentivising research that maximises an individual’s performance based on a particular metric or set of metrics.**

This phenomenon exemplifies Goodhart’s Law: that is, statistically speaking, once a metric becomes a goal, it ceases to be a valid measure of what it was intended to measure. Once quantitative metrics are instituted then there is incentive to maximize one’s ranking. Such incentives change research behaviour in an undesirable way, moving it in directions that maximize the metric performance and not necessarily the needs of society or the curiosity-driven interests of a researcher.

Additionally, when metrics aim to indicate ‘impact’, (e.g., publication in international journals, citations) then the incentive becomes tied to doing research on high profile topics that will be published in high visibility journals to indicate impact, as per the metric. Judging research quality in this way ignores the fact that important and high-quality research is needed throughout society, not just on flashy and international issues or regions. The quality of this sort of research cannot be judged by ‘visibility’ metrics. This is an undesirable outcome because research is largely locally funded (provincially and nationally) and setting goals that incentivizes non-local research creates a system where we as a society are often paying salaries and funding research on the issues that most directly benefits others.⁸

This phenomenon was discussed in Yves Gingras’ book when he presented a scenario in which Canadian academic economists may tend to disproportionately study American and global economies at the expense of research on Canadian issues. This was due to striving for high visibility research and to the detriment of the society that funded the research”

“An Economist who wants to “maximize” its number of citations would thus tend to study the economy of the United States rather than that of France or Canada, which are of little interest to American journals of economics that happen to be the most cited.”⁹

7. **Numerical components of bibliometrics are gender-biased.**

All bibliometrics use some combination of the number of papers, the number of citations, or authorship order in their calculations. Larivière et al. summarized all research included in ISI’s Web of Science database (>5 million records 2008 - 2013) across all disciplines.¹⁰ The analysis demonstrated that in all disciplines and all localities, men published more papers than women, women held fewer senior authorship positions on papers than men, and women were cited less often. This work was a summary of the raw data on publication practices and so this bias is inherent to all bibliometrics that rely on these values. The committee agreed that systematic discrimination of any sort in evaluation is unacceptable.

---

8. **Databases used to calculate bibliometrics are not well-correlated at the individual level.**

This point was elucidated in a seminar presented by Vincent Larivière at the University of Manitoba on November 23, 2017.\(^\text{11}\) Bibliometrics are reliably calculable when the same database is used. However, database information is not well correlated when correlations are assessed at the individual level. This lack of correlation makes the use of metrics for the assessment of individuals lack robustness: they vary considerably from one database to the next and there is no objective basis for deciding which database is best or provides greater accuracy.

9. **There are currently no metrics that suit the criteria for usability in the evaluation of individuals.**

Despite an effort to uncover metrics that met the committee’s criteria, it was agreed, after extensive research and consultation with the scholarly literature, that there are no current bibliometrics suitable for the evaluation of individual research quality. There was clear utility for the use of metrics to assess ‘populations’ of individuals. However, the use of these metrics for assessing individuals suffered from the textbook statistical and logical fallacy termed the ‘ecological fallacy.’ This well-characterized statistical logical fallacy occurs during the interpretation of statistical results and states that one cannot draw inferences about the nature of individuals from inferences about the group to which an individual is a member. While the utility of bibliometrics at the group level was in some cases clear, the Committee was not able to identify a viable or valid way for group level indicators to be translated to individuals level indicators of research quality.

Recommendations

In light of the numerous flaws and issues presently associated with the application of research metrics in the evaluation of scholarly work, the Joint Committee agreed on a set of principles and guidelines that are necessary in order for a metric to be considered valid.

Much of the Committee's work in this area was informed by Gingras' *Essential Characteristics of a Good Indicator.*

Principles of Evaluation for Bibliometrics

1. Peer review is foundational.
2. Bibliometrics should strive for objectivity.
3. Bibliometrics should be customized to the discipline.
4. Bibliometrics should minimize bias.
5. Bibliometrics should maximize objectivity.

Guidelines in the Use of Bibliometrics

1. Bibliometrics that are flawed should be avoided.
2. Bibliometrics should be empirically validated measures related to well-defined features of research or researcher quality.
3. Bibliometrics should be field-weighted.

---

Selected Bibliography on Bibliometrics


64. Hjørland B. Citation analysis: A social and dynamic approach to knowledge organization. Information Processing & Management. 2013;49(6):1313–25.


100. Schoonbaert D, Roelants G. Citation analysis for measuring the value of scientific publications: quality assessment tool or comedy of errors? Tropical Medicine & International Health. 1996;1(6):739–52.


LETTER OF UNDERSTANDING

Re: Joint Committee on Metrics

BETWEEN:

THE UNIVERSITY OF MANITOBA (Hereinafter referred to as “the University”)

-and-

THE UNIVERSITY OF MANITOBA FACULTY ASSOCIATION (Hereinafter referred to as “UMFA”)

1. The University of Manitoba believes that, in light of new provisions in the collective agreement enhancing collegial participation in the creation of tenure and promotion criteria, Members should have the ability to decide on the appropriate use (if any) of research metrics (as defined in s. 19.D.1.8.1) in tenure, promotion, and performance evaluation processes.

2. The University of Manitoba Faculty Association believes that any such requirements pose risks of unfair, inaccurate, and/or discriminatory processes of assessment in tenure, promotion, and performance review. It further believes that such requirements pose significant risks to the quality and integrity of research, scholarship, and education. It believes that therefore, no such requirement should be imposed on Members, nor should Members be penalized for refusing to submit such metrics as part of tenure, promotion, or performance review.

3. The University of Manitoba and the University of Manitoba Faculty Association agree that a Joint Committee will be struck to examine issues related to the collection and use of research metrics in evaluative processes (such as tenure, promotion, and performance review).

4. The University and the Association will each nominate three members (each with academic rank below the level of a Dean, and who are not on the UMFA executive) to the committee, including one to act as co-chair.

5. In the course of conducting its work, the Joint Committee will:

   a) examine the academic and scholarly literature relevant to the appropriateness, risks and benefits of the collection and use of research metrics as a means of assessment;
b) examine the treatment of research metrics at other Canadian Universities;

c) consult with members of the academic community representing a diversity of disciplines;

d) consult with members of the academic community with experience and/or relevant academic expertise relating to the collection and use of research metrics.

6. The Joint Committee will submit a report to the President of the University and the President of the Association containing its findings and recommendations no later than December 31, 2017.

7. Should the committee recommend it, with no more than two (2) members dissenting, the University and the Association agree to the immediate addition of the following to the Collective Agreement:

19.D.1.8.5 No tenure criteria shall require the submission of a research metric, nor shall they include any standard or recommended expectation based on a quantitative measure.

19.D.1.8.6 Research metrics will only be used as part of evaluation and/or assessment when such metrics are personally and voluntarily submitted by the Member.

19.D.1.8.7 There shall be no penalty or adverse inference to any Member for the non-submission of research metrics as part of the tenure application.

20.A.2.5.5 No promotion criteria shall require the submission of a research metric, nor shall they include any standard or recommended expectation based on a quantitative measure.

20.A.2.5.6 Research metrics will only be used as part of evaluation and/or assessment when such metrics are personally and voluntarily submitted by the Member.

20.A.2.5.7 There shall be no penalty or adverse inference to any Member for the non-submission of research metrics as part of the promotion application.

20.B.1.6.5 No promotion criteria shall require the submission of a research metric, nor shall they include any standard or recommended expectation based on a quantitative measure.

20.B.1.6.6 Research metrics will only be used as part of evaluation and/or assessment when such metrics are personally and voluntarily submitted by the Member.

20.B.1.6.7 There shall be no penalty or adverse inference to any Member for the non-submission of research metrics as part of the promotion application.
35.5.5 No evaluation criteria shall require the submission of a research metric, nor shall evaluation criteria include any standard or recommended expectation based on a quantitative measure.

35.5.6 Research metrics will only be used as part of evaluation and/or assessment when such metrics are personally and voluntarily submitted by the Member.

35.5.7 There shall be no penalty or adverse inference to any Member for the non-submission of research metrics as part of an evaluation.

8. Notwithstanding s. 7, the Joint Committee shall be free to come to any other recommendations or conclusions it desires, including alternative wording for the principles described in s. 7. Any recommendations under this section will be advisory only.

9. This letter of understanding shall be attached to and form part of the Collective Agreement.

10. This letter of understanding will expire on January 12, 2018. This letter of understanding shall be attached to and form part of the Collective Agreement and the period in which s. 10(4) of the Labour Relations Act applies.

11. The terms of this letter of understanding are a term and condition of employment. This letter of Understanding is intended to survive the expiry of the Collective Agreement.

DATED at the City of Winnipeg in the Province of Manitoba

this _____ day of ________ 2017

____________________________________
University of Manitoba

____________________________________
University of Manitoba Faculty Association