Citation Analysis Methodology

The Working WISE citation analysis was conducted to identify what scholars have been writing about workplace barriers for women in the science, technology, engineering and mathematics (STEM) fields and which of their articles have been most frequently cited by others. The main goals were to uncover the work that has been most influential in this area of research and to trace research themes over time. We were most interested in research articles that have appeared in scholarly outlets and that directly address the issue of work barriers. We began the literature search and citation analysis by establishing some initial parameters that we felt would focus the analysis on our particular goals. Thus we decided to include articles that 1) were published from 1970 to the present, 2) were peer reviewed, and 3) address workplace factors that influence the recruitment, retention and promotion of women in STEM fields. We decided to exclude articles which were primarily about the educational system and pipeline issues. These topics were being actively addressed by others and thus were not the primary focus of the Working WISE Project.

Phase 1

As we began to search for articles, we found we needed to establish several other parameters and guidelines. We decided to include articles from non-refereed magazines in the STEM fields if they were substantial (i.e., 3-4 pages or more) since many timely articles are published in these field-specific outlets. Although pipeline issues are not directly within our charge, we decided to include university-based recruitment programs when they were related to encouraging women to stay in STEM fields and when they were substantial and research-based. Given that we were not initially finding as many articles about specific science fields, we decided to specify fields (e.g., mathematics, physics, chemistry, biology) when doing searches. We used APA format for all citations and bibliographies.

We also made some decisions about what types of publications and topics to exclude in order to keep the citation analysis focused. In the initial searching we discovered numerous dissertations in our general subject area, some of which focused on women working in business settings. We made the decision to exclude the dissertations, however, because the process of culling the relevant ones from those that were not relevant would have been extremely time intensive, as a large number appeared to be extremely narrowly focused on a particular industry, location or time period. We reasoned that the research results that fit our criteria most closely would emerge in subsequent published studies.

We also decided to exclude the medical and nursing fields for this initial search. This decision was shaped by several issues. Most importantly, the issues for women in medicine tend to be different from the concerns for women in specific STEM fields, e.g., women are more evenly represented in medical training programs even though they tend to go into different subspecialties. Nursing is an overwhelmingly female-dominated field, which presents a whole different set of issues. These circumstances raise a series of complex sub-questions that are currently outside our scope. Additionally, the research on women in medicine and nursing fields tends to be a separate literature and is not typically addressed in the same articles as those about women in STEM fields. If we find we have time, we may address the medical field in a subsequent analysis.

Our goal was to come up with three end products: 1) a list of the most prolific authors, 2) a list of the most highly cited papers, and 3) an examination of trends over time (comparing articles grouped by 5-year periods).

We utilized *RefWorks* bibliographic management software to manage the search. *Ref Works* allows automatic downloads from many library databases and can produce bibliographies in multiple formats, as well as generate proper citations or footnotes for manuscripts. To track patterns in citations over time, we developed electronic folders for articles published in 5-year groups.

Using the established parameters, we searched numerous databases, including:

1. Ebsco (vendor): searched in Academic Search Premier, Sociological Abstracts, PyscInfo/PyscArticles, ERIC (Education), EJS table of contents, Business Source Premier.
2. ProQuest (vendor): checked both Academic Journals and Trade Publications, ignored Newspapers and Magazines.
3. Infotrac (vendor): Expanded Academic Index, General Reference Center Gold, General Business File
4. Compendex: Engineering Index
5. ScienceDirect
6. CSA (vendor): used Sociology and Biological Sciences databases.
7. IEEE: electrical and computer engineering
8. World Cat Searching for books
9. Math Reviews
10. MultiSearch: tried with the Sciences set, helpful but time consuming due to limitations in the Multi-Search platform. Abandoned in favor of searching individual vendors.
11. Ingenta, a table of contents service.

In general, the search was done with up to four sets of keywords. If a database yielded a small number of total hits with only two sets, we did not add additional keywords. The keyword sets, ranked in order of use, were as listed below. (Please note that * is the usual truncation symbol.)

1. wom*n or minorit* or female*
2. science or engineer* or math* which was also modified to specific science fields (biolog* or physics or chemist*) in certain databases
3. career* or job* or employ*
4. retention or retain* or promot* or advance* or recruit*

This process yielded a large number of items, estimated to be about 1400 of the 2,009 articles which ended up in the RefWorks database. The total number of articles in the RefWorks database reflected items downloaded from more than one database, as many items were found in more than one place. As a general rule during this process, citations were downloaded without checking to see if they were already present, because RefWorks contains an excellent duplicate detection system which was heavily utilized by the researchers. Using this process, 973 items downloaded into RefWorks were subsequently removed either because the item was duplicated, or because it was later judged by the Principal Investigators (PIs) to be outside the scope of the study. The final total number of identified articles was 1,036.

As citations began to accumulate, the Principal Investigators reviewed them and designated "Key Articles." We began by excluding articles that were not as directly related to barriers for women working in STEM workplaces in the United States. For example, we excluded articles that addressed:

1) General issues facing women and work which were not specific to STEM fields;
2) Education and career choices, particularly if the focus was on K-12;
3) Individual's stories, which are often biographical or autobiographical and tend to be too anecdotal;
4) Concerns in other countries unless there was some comparison to issues in the U.S.;
5) Descriptions only of how many women work in a particular science field.

The positive criteria for selection of articles determined to be "key articles" were number of times cited, national reputation of scholar, seminal articles considered to be "classics," representation across fields, years of publication, and themes.

Phase 2

As a result of the literature review process, 285 were designated as "key articles." We then researched how often each of these articles had been cited in the literature. To do this, it was necessary to use Web of Science (WofS), a product of the Institute of Scientific Information, recognized world-wide as the premier citation information database. It tracks and counts the number of times an article has been cited by every article in its extensive database of over 12,000 journals, going back to the 1940's. The Boston University license accessed all articles written from 1983 forward. Articles written before 1983 could still be checked via cited reference search to see how many times they had been cited in the literature from 1983 to the present. However, a number of articles designated as KEY were not in WofS at all, and some, while present in WofS, had not been cited by anyone. Any articles not found in WofS via author search were searched carefully as a cited reference. This turned up a number of citations.

WofS allows several different types of searches. The easiest one, called a "General Search," was done first for all items, in which each citation was simply searched in the database by the easiest method. If an item was in WofS, the results of the search showed how many times other authors had cited the item. Usually searching by the author's name, crossed by year of publication, retrieved the desired result. A second search was performed if no results were found from a General Search, using the "Cited Reference" search. In this search, a different section of the database was accessed, that is, all the citations from all the articles in the database. Searching by author's last name and crossing by year produced a long list. This list was then scrutinized for the correct article, and the number of citations was recorded. In cited reference searching, multiple forms of the author and work were encountered, as WofS makes no attempt to fix incorrect citations.

While doing author searches of our key authors, other articles of interest kept appearing. These articles were downloaded and offered to the PI's to assess whether they should be included as Key Articles. Because of possible interest the number of citations to these new articles was carefully recorded.

Books had to be searched differently from articles. Forty-two books or book chapters had been designated as "Key" and needed to be searched in WofS. All book chapters were searched carefully and completely in WofS, as items of interest could be cited as a page number, a chapter number or as an independent essay in a book. An additional spreadsheet for the book list was designed, as well as one for lists of items Not Found and Not Cited. Author counts were taken directly from RefWorks, which includes an easy and comprehensive way to see how many times a particular author was present in the database. It had to be checked carefully because each version of an author's name was recorded on a separate line. Therefore, authors could be listed with full

names, with initials only, and with or without middle initials. Student research assistants added up the variant forms and recorded the author totals. The RefWorks file of all items was cleaned up, so that only folders of Key articles, Books, Not Found/Not Cited and year spans were left in the file. A comprehensive list in bibliographic format was printed out.

The larger database of 1,036 items (86 Books or book chapters, 123 Conference papers, 13 reports, and 814 journal articles) were looked at in 5 year periods to determine the trends of each period, the main authors, and other such tends. The lists compiled were:

1) Times Cited—the number of times each of the Key articles was cited in the literature. This covered the 87 designated key articles.
2) Authors—the number of articles each author had written or co-authored since 1970. This list encompassed the entire 1,036 item database.
3) 42 Books—The number of times each book was cited in the literature
4) 77 Cited but not KEY—Other articles in the database which were cited by others but were not considered KEY articles. Created in case needed later, as the information had been collected and was kept available.
5) 123 Not found/Not cited—articles originally considered KEY but either not found in WofS or not cited by anyone. These articles were removed from the list of Key articles. Many of them were from the non-refereed journals.
6) 31 Not Cited in 2004-2005. This special group consisted of recent articles which were less likely to have been cited by others because they were newer. If they had been cited, the number would be low because of the recent publishing date. These items would still be considered KEY if the PI's designated them as such.

The Key articles were further cut back from 87 to 46, then accessed, read and reviewed by the Working WISE Project Manager and Research Assistant. Criteria for inclusion in this group included:

a. Number of times cited;
b. Representation of important authors in the field;
c. Inclusion of authors from each generation;
d. Principal Investigator's knowledge of the literature and the field;
e. Inclusion of articles on each STEM field.

The literature reviews were passed on to the PI's to use in structuring the Delphi process.

Marion Muskiewicz,
Lydon Library
University of Massachusetts Lowell
August 15, 2006