

Citation Analysis

Do Age and Types of Materials Cited Correlate with Availability of Appropriate Library of Congress Subject Headings?

C. Rockelle Strader

This study is a citation analysis of a set of theses and dissertations in the Ohio State University's online catalog, for which the author-assigned keywords and cataloger-assigned Library of Congress Subject Headings (LCSH) are known. Correlations are sought between the types and ages of resources cited and the number of unique keywords and unique LCSH that were found. The author presents results found in three general discipline areas: arts and humanities, the social sciences, and science, technology, engineering, and medicine.

The process of creating new Library of Congress subject headings (LCSH) can be lengthy and consequently lag behind common use of terms in research. New subject headings are proposed and established using literary warrant when a cataloger is cataloging an item and is not satisfied with the available LCSH. Subject cataloging requires consultation of reference works—to the extent that relevant sources are available—to determine concepts and terminology that prevail at the time the heading is created. Does the currency of subject headings correlate with specific disciplines or larger fields of study? Does their availability correlate with the type of literature published in a field or with the type of literature most often cited in that field's literature?

The research reported here explores these issues through a citation analysis of a set of electronic theses and dissertations (ETDS) with author-assigned keywords and cataloger-assigned subject headings. The author sought to answer the following research questions:

- Do broad disciplines, i.e., arts and humanities; social sciences; and sciences, technology, engineering, and medicine (STEM) evidence-specific citation patterns? That is, do ETD authors in these fields cite a specific type of material, e.g., monographs, journal articles, maps, or websites more or less frequently?
- What patterns within disciplines can be identified when type and age of cited materials are examined?
- What correlations exist between type of cited materials and unique keywords or LCSH?
- What correlations exist between age of cited material and unique keywords or LCSH?

C. Rockelle Strader (strader.2@osu.edu) is Catalog Librarian at The Ohio State University Libraries, Columbus, Ohio.

Submitted December 31, 2011; returned to author February 2, 2012 with request to revise and resubmit; revision submitted April 3, 2012, reviewed, and tentatively accepted May 12, 2012, pending modest revision; revision submitted June 10, 2012, and accepted for publication.

- Do these findings have implications for the maintenance of LCSH?

The author tested two assumptions:

- The arts and humanities will make heavier use of monographs, STEM will have very high use of serial articles and very little use of monographs, and the social sciences will fall between arts and humanities and STEM.
- Research in disciplines that cite monographs more often will show fewer unique keywords and LCSH.

The project described here builds on the author's 2009 study, "Author-Assigned Keywords versus Library of Congress Subject Headings."¹ In that study, keywords assigned by ETD authors were compared to LCSH assigned by catalogers when describing the same ETD. The purpose was to determine the percentage of keywords that were unique and could provide useful points of entry into the catalog. The author examined six categories of match: exact match, all present (but not in exact order), partial match, needs two LCSH, variant, and no match. The results for keywords found that 25.16 percent of the time they matched LCSH exactly; all present (but not in exact order), 3.39 percent; needing two LCSH, 2.26 percent; and partially, 24.69 percent. Keywords that were variants of LCSH appeared 9.93 percent of the time, and did not match at all 34.56 percent of the time. Conversely, LCSH matched keywords exactly 36.49 percent of the time; all present (but not in exact order), 4.49 percent; and partially, 31.06 percent. LCSH as variants of keywords appeared 11.34 percent of the time, and did not match at all 16.60 percent.

In the previous study, the author noted that

one explanation for the large percentage of terms not covered by cataloger-assigned LCSH is that LCSH has not kept up with current research. . . . LCSH typically are established from evidence of a new topic found in the piece in hand, that is, from literary warrant. This is usually a monograph in hand, since articles and chapters are generally not cataloged. However, in some disciplines, such as the physical sciences and medicine, the journal literature is the primary publication environment for new research, and dissertations in those fields could be among the first comprehensive monographic treatments of a topic that has been otherwise extensively discussed.²

In the present study, the author explored the types and characteristics of literature using the previous study's population of 285 electronic theses and dissertations (ETDs) for

citation analysis, looking for patterns that correlate with availability of LCSH. Collectively, the ETDs contain more than 38,000 citations. The author examined these citations for patterns in the age and type of materials cited (book, journal article, etc.) within the three broad discipline areas of the arts and humanities, the social sciences, and STEM. This study is intended to contribute to the literature on the use of materials in graduate-level research and the patterns of use within disciplines, as well as to the literature on the use LCSH for access to these resources.

LCSH and Literary Warrant

Warrant is a foundational concept that underlies the presentation and searchability of terms in systems like library catalogs and databases. Using warrant to select terms for a controlled vocabulary is explained in *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*, which states,

The process of selecting terms for inclusion in controlled vocabularies involves consulting various sources of words and phrases as well as criteria based on:

- The natural language used to describe content objects (literary warrant)
- The language of users (user warrant), and
- The needs and priorities of the organization (organizational warrant).³

The standard goes on to state that "assessing literary warrant involves reviewing the primary or secondary content objects that the vocabulary will be used to index as well as consulting reference sources such as dictionaries or textbooks and existing vocabularies for the content domain."⁴ This paper focuses on literary warrant. User warrant addresses how users choose terms for searching, which is outside the scope of the current study, as is organizational warrant.

Literary warrant is the foundation of the LCSH system. The initial and still-primary purpose of LCSH is to be a controlled vocabulary for the Library of Congress (LC).

The Library of Congress collections serve as the literary warrant (i.e., the literature on which the controlled vocabulary is based) for the Library of Congress subject headings system. The number and specificity of subject headings included in the Subject Authority File . . . are determined by the nature and scope of the Library of Congress collections. Subject headings are established as they are needed to catalog the materials being added to

the collection or to establish links among existing headings. In recent years, headings contributed by libraries engaged in cooperative activities with the Library of Congress based on the needs of their collections have also been included.⁵

Shared cataloging and initiatives such as the Program for Cooperative Cataloging (PCC) have expanded the scope beyond the collection of the LC. LCSH is not intended to be a comprehensive thesaurus; rather, it is for the assignment of subject access points on an as-needed basis.⁶ Consequently, creation of new subject headings will to some extent always lag behind the publication of works in any given discipline.

Lag time and scope are criticisms frequently leveled against LCSH. For example, in her discussion on the complementary nature of keywords and LCSH in cataloging and searching, McCutcheon noted that the weaknesses of LCSH are closely related to its strengths.

The very things that make LCSH such a powerful finding tool—its consistency, accuracy, and depth—also make it relatively slow to adopt new terms. New subject headings are proposed by catalogers based on literary warrant—that is, a work must be written then cataloged using available, inadequate terms; the new subject heading is submitted and reviewed before it is approved for use in the thesaurus. . . . The procedures in place to ensure quality control result in a significant delay between the point when a cutting-edge discovery makes its debut and the time when terms conveying the nature of that discovery enter the LCSH canon.⁷

McCutcheon posited that the use of keywords in combination with LCSH mitigates the lack of the most current terminology in the controlled vocabulary. Utilizing both maximizes the chances of successful discovery, particularly with resources representing current research, such as ETDs. McCutcheon summarized, in the context of ETDs at one university, a problem that is recognized by the LC. The Cataloging Policy and Support Office's report, "Library of Congress Subject Headings: Pre- vs. Post-Coordination and Related Issues," noted that, while literary warrant reflects the topics of materials that have been acquired, it is "criticized for lacking a universal knowledge structure to anticipate topics or for lacking a way for users to easily contribute to the terminology."⁸

The LC *Subject Headings Manual* states that a new subject heading may be proposed when it "represents a discrete, identifiable concept when it is first encountered in a work being cataloged."⁹ However, the topic of a new work being cataloged may be judged not sufficiently discrete or identifiable, therefore a new subject heading is not advised.

Further, even when a topic is deemed discrete and identifiable, consensus within the discipline regarding the "proper terminology for the concept" may be absent.¹⁰ If this is the case, the cataloger must engage in authority research to decide on the term that best suits the title in hand. The LC guidelines suggest that very new topics may not be perceived as sufficiently settled to warrant their own headings at the time they are first encountered.

Literature Review

This review focuses on the literature reporting citation analysis research. Every discipline of study has its own history and particular terminology, which must be learned and appropriately used by new and emerging scholars for further dissemination of knowledge. A common practice is to place new research in the context of previously published domain-related works. One way to examine the context into which a given work falls is to conduct a citation analysis, which examines the frequency and patterns of citations in a representative sample.

The types of resources cited in documents have been shown in citation analyses to be a reflection of general discipline area. Many of these studies generalize into three broad groups: arts and humanities, the social sciences, and STEM. For example, Nederhof conducted a study on publication patterns and citation behavior in the "(basic) sciences and the social sciences and humanities."¹¹ His results pointed to the tendency of the humanities and social sciences to publish more in books than do the sciences. Similarly, Huang and Chang studied various characteristics of research output in the social sciences and the humanities.¹² They documented a wide range and diversity of publication patterns by these two groups, with significantly more publication in and citation of books and book chapters than found in the sciences.

Cullars found a heavy reliance on monographs and manuscripts in his study of fine arts monographs.¹³ He further noted that although the reliance on monographs was generally consistent with previous studies in the humanities, his findings were slightly less, while the use of manuscripts in his sample was greater than the rates reported in other studies. Dealing with the other two discipline areas, Glänzel and Schoepflin examined differences between the sciences and the social sciences in the number of references to serial and nonserial literature.¹⁴ They demonstrated that the sciences are much more reliant on journal articles than the social sciences.

Citation analyses that are limited to theses and dissertations generally find the same patterns in discipline areas. Buchanan and Herubel studied a small sample of philosophy and political sciences dissertations, representing the humanities and the social sciences, respectively.¹⁵ While

both disciplines exhibited more reliance on monographs than journals, the authors found heavier use of monographs in the philosophy dissertations while the political science dissertations cited more out of the journal literature.

Kushkowski, Parsons, and Wiese conducted a longitudinal study of masters and doctoral theses publishing during a twenty year period and discovered a preference for the journal literature in engineering and the sciences.¹⁶ The arts and humanities, on the other hand, cited more monographs, although the journal citation rate for that group climbed significantly in the last five years of the study period. The social sciences exhibited a preference for the journal literature but had a noticeably higher rate of citation to other theses.

Some citation studies of theses and dissertations have found varying results. For example, Wehmeyer and Wehmeyer found a higher-than-expected reliance on monographs in clinical psychology dissertations.¹⁷ Other studies, such as that by Conkling and colleagues, found a growing reliance on the journal literature across disciplines.¹⁸

In addition to format, the age of cited materials also varies by discipline. Resources, both older and current, provide the background and context for the concepts and terminology used in a given document. As Rogers noted, the use of terms changes over time as concepts are introduced and subsequently developed, as new information and knowledge supplant and make obsolete older ideas and theories, and as social and cultural conventions change, and so on.¹⁹

Glänzel and Schoepflin examined the use of serial versus nonserial resources and sought to determine the relation between age of references and their format and the possible correlation between “reference age and share of (non-)serial literature.”²⁰ While they found significant differences in the use of serial resources, their results for the age of resources used was mixed. In Nederhof’s research, sources cited in biomedical fields showed a mean age of between seven and eight years, while history and philosophy of science and social sciences had a mean age of thirty-nine years; however, the mean ages of citations used in the fields of physics, mathematics, psychology, and economics all fell roughly between ten to eleven and one half years.²¹ Huang and Chang found a higher tendency to use literature at least ten years old in the social sciences and humanities.²² They determined that the social sciences and humanities tend to cite more books than the sciences do, and concluded that these contribute to a wider age distribution of cited literature. Cullars also found a wide age distribution in his study of fine arts monographs.²³ While only 10.8 percent of the cited resources were published in the twelve years preceding his research, 47.3 percent of the cited resources were published in the previous thirty-two years; the rest were older, with 18.6 percent published before 1890.

Buchanan and Herubel also found variation in the age of citations in their study.²⁴ While the philosophy and political

science dissertations they examined more frequently cited monographs, the publication date patterns were markedly different for monograph and journal citations. Only 17 percent of monographs cited in philosophy were published in the five years preceding their research compared to 34 percent in political science. However, the journal articles cited in both samples had similar characteristics. About 51 percent of journal articles cited in philosophy and 55 percent cited in political science were published in the previous five years.

Hargens found that two basic citation patterns emerged depending on discipline: some disciplines focused on and cited recent research while ignoring foundational work; conversely, some focused on early work while ignoring recent publications.²⁵ Kushkowski, Parsons, and Wiese attempted to corroborate Hargens’ results, but instead found that dissertation authors across all disciplines tended to prefer current research, a preference that increased over the course of their longitudinal study.²⁶ Arts and humanities dissertations showed the most variability in the age of their citations. At the beginning of the twenty-year sample, the mean age of citations was more than thirty-five years while the mean age of citations in dissertations written in the most recent five years fell to just under ten years. This appeared to correlate with the increase in journal use in that group during the same period.

Conkling and colleagues found variations in the age of materials cited in their study of pre-web (1990s) and post-web (2000s) dissertations in the social sciences, sciences, and engineering.²⁷ Journal use increased, monograph use decreased, and the age of cited materials trended toward slightly older materials in dissertations written after the advent of the web.

Research Method

The author collected the data for this study from bibliographies of the 285 ETDs used in her 2009 study and from the keyword and LCSH data gathered at that time.²⁸ These ETDs were submitted to the OhioLINK ETD Center by doctoral candidates between June 1 and October 31, 2005, and Ohio State University catalogers received automatic e-mail notifications of their availability during the same period. To be considered part of the population studied, the ETDs had to have author-assigned keywords and full text available at the time of cataloging and to be cataloged by the end of 2006.

The author entered citations from each ETD bibliography in a Microsoft Excel spreadsheet and coded them for type of material and year of publication or completion, presentation, or availability in cases such as preprints, uncollected conference papers, or unpublished theses and dissertations. The author assigned eight material types: “book chapters”; “media, maps, and music”; “monographs”;

“proceedings and presentations”; “serial articles”; “theses”; “webpages and websites”; and other. Monographs included print and electronic monographs. “Serial articles” included those published in print or online professional journals and magazines, newsletters, popular periodicals, newspapers, and newswires. “Webpages and websites” included online content that could not be characterized as an electronic monograph or serial article. “Other” included archival materials, correspondence, course materials, data sets, National Science Foundation proposals, patents, press releases, software, testimonies from hearings and trials, and those that the author was unable to determine.

The author also coded citations for discipline area on the basis of the major or academic unit (which could be a graduate program, department, school, or college) as identified by the ETD’s author on the title page. Separation of majors into the three discipline groups was based on the Ohio State University colleges that contained the majors and academic units in 2005 (the year in which all the ETDs were submitted). In 2005, majors in the arts and humanities were located in two colleges: Arts or Humanities. Majors in the social sciences were located in five colleges: Business; Education; Human Ecology; Social and Behavioral Sciences; or Social Work. STEM majors were located in nine colleges: Biological Sciences; Engineering; Food, Agricultural, and Environmental Sciences; Mathematical and Physical Sciences; Medicine; Nursing; Pharmacy; Public Health; or Veterinary Medicine (see appendix).

The initial collection consisted of 38,424 citations. Citations that were duplicated within a given document were removed, leaving 38,317 citations for discussing material type. Citations to a resource that occurred in more than one ETD were each kept because they form the basis of each ETD author’s use of terms. The author removed an additional 230 citations from the sample when determining the age characteristics of the materials used (for a total of 38,087 citations) because their ages could not be deduced from the citations provided. The author calculated average and median ages of citations (from which the age of the citation could be determined) for each ETD by material type and across all material types.

To address the questions of correlation between material type, age of cited works, and assignment of keywords and LCSH, the author used data from the previous study. These data were the percentages of author-assigned keywords for which no corresponding cataloger-assigned subject headings (“unique keywords”) were present (34.56 percent), as well as the percentages of LCSH assigned by catalogers for which no corresponding author-assigned keywords (“unique LCSH”) existed (16.60 percent).²⁹

Data on variants of keywords and LCSH also were taken from the previous study. Although, in that study, variants were treated as a form of match and thus not part of

the count for “unique” keywords and LCSH, they still may be considered to be nonmatches, with implications for the maintenance of cross-references on which LCSH depends. In the previous study, keyword variants came in three forms: different word forms and spellings found in LCSH, variant forms and spellings of LCSH cross-references, or abbreviations of LCSH terms (such as chemical symbols). Collectively, keyword variants composed 9.93 percent of the keyword matches against LCSH. LCSH variants also had three forms: those instances in which the established terms were (or contained) variant forms of keywords, instances in which the cross-references were (or contained) variants of keywords, or instances in which the LCSH were (or contained) abbreviations of keywords. Collectively, LCSH variants were 11.34 percent of the LCSH matches to keywords. Unique keywords and keyword variants taken together made up 44.49 percent of the matches against LCSH, while unique LCSH and LCSH variants together made up 27.34 percent of the matches against keywords.

Linear correlation (p value = 0.05) was performed in two sets: (1) between material type and unique keywords and unique LCSH, as well as keyword and LCSH variants ($N = 38,317$); and (2) between median age of each material type and unique keywords and unique LCSH, and keyword and LCSH variants ($N = 38,087$). In each set, the correlations were run under four variable conditions each for keywords and LCSH: (1) unique keywords and LCSH separately; (2) each type of keyword and LCSH variant separately; (3) all keyword variant types together and all LCSH variants together; and (4) unique keywords plus all keyword variants together (hereafter, “all keyword non-matches”), and unique LCSH plus LCSH variants together (“all LCSH non-matches”).

Correlation is measured by the correlation coefficient, which indicates the strength of the correlation and whether the correlated variables move together in the same direction (positive) or opposite each other (negative). The closer the correlation coefficient is to 1 or -1, the stronger the correlation (positive or negative, respectively) between two variables. The p value indicates the statistical significance of the correlation and for this study was established at $p = 0.05$ or less. That is, for p -values of 0.05 or less, the likelihood is lower that the result was produced by chance.

Results and Discussion

Initial analysis sought to answer the first two research questions:

- Do broad disciplines, i.e., arts and humanities, social sciences, and STEM, evidence specific citation patterns?

Table 1. Citation Material Types (*N* = 38,317)

Material Types	Arts and Humanities		Social Sciences		STEM		Total	
	Number	Percent	Number	Percent	Number	Percent	Number	Percent
Book chapters	440	11.01	1,743	11.90	749	3.81	2,932	7.65
Media, maps, music	71	1.78	3	0.02	1	0.00	75	0.20
Monographs	2,053	51.39	3,526	24.08	1,939	9.85	7,518	19.62
Other	63	1.58	145	0.99	245	1.24	453	1.18
Proceedings, presentations	136	3.40	336	2.29	1501	7.63	1973	5.15
Serial articles	986	24.68	8,548	58.38	1,4931	75.87	24,465	63.85
Theses	157	3.93	214	1.46	161	0.82	532	1.39
Webpages, websites	89	2.23	127	0.87	153	0.780	369	0.96
Total	3,995	100.00	14,642	99.99*	19,680	100.00	38,317	100.00

*Does not equal 100% because of rounding.

- What patterns within disciplines can be identified when type and age of cite materials are examined?

Discipline and Material Type

Most materials cited in the ETDs were serial articles (63.85 percent of all citations); see table 1. Monographic works were used far less frequently (19.62 percent). However, significantly different patterns emerge when the data are examined in the context of the discipline areas. The assumption that the arts and humanities would make heavier use of monographs, that STEM would have very high use of serial articles and very low use of monographs, and that the social sciences would fall between arts and humanities and STEM, was confirmed. In the arts and humanities, 24.68 percent of the citations were for serial articles and 51.39 percent were for monographic materials. STEM citations were radically different, with 75.87 percent of the citations to serials and only 9.85 percent to monographs. In the social sciences, 58.38 percent were to serial literature and 24.08 percent were to monographic literature.

Discipline and Age of Cited Materials

Examining the ages of cited materials revealed similarities in the arts and humanities in contrast to STEM (with social sciences in the middle); see table 2. Cited materials were, on average, older in the arts and humanities than in the social sciences or STEM. This pattern held for median age as well. Although the median age of the sources cited in arts and humanities may seem to be young at thirteen years, it is still older than the median eight and nine years for STEM and social sciences, respectively. This finding confirms previous studies, which also found that the arts and humanities tend to use older materials. Median age was calculated to more

accurately gauge where most of the materials clustered by age, thereby reducing the effects of outliers on the statistics. In this study, the outliers were significantly older, with the oldest ranging from 115 to 406 years. The oldest cited document was a monograph used in the arts and humanities group.³⁰

A related pattern that emerged is that the average and median ages of cited monographs tended to be older than cited serial articles. This pattern held across all three discipline groups. However, the oldest cited material in the STEM ETDs was a serial article, which, at 216 years, is older than the oldest cited serial article in the arts and humanities, which was 116.

Three research questions remained:

- What correlations exist between type of material cited and unique keywords and assigned LCSH?
- What correlations exist between age of material and unique keywords and assigned LCSH?
- Do these findings have implications for the maintenance of LCSH?

Material Type and Unique Keywords and LCSH

Table 3 summarizes the correlations found between material type and the keyword and LCSH variables (unique keywords and LCSH, keywords and LCSH variants, all nonmatches). It collocates all the results first by material type, then by discipline area, and then further by keyword and LCSH variable. The correlation coefficient and p value (strength and direction of correlation and statistical significance) are shown next to each variable. The most notable result of this part of the study was that no statistically significant correlations were found between any material type and unique LCSH. Further, no significant correlations were found for the material types of “media, maps, and music.”

Table 2. Age Characteristics by Discipline Group (N = 38,087)

Material Type	Arts and Humanities	Social Sciences	STEM	Overall
	Average age in years			
All materials	22.89	13.17	11.40	13.25
Book chapters	12.90	12.10	11.36	12.03
Media, maps, music	44.81	5	23	42.90
Monographs	25.64	15.73	14.96	18.23
Other	8.87	9.06	7.91	8.40
Proceedings, presentations	10.80	9.40	8.25	8.62
Serial articles	24.22	12.73	11.38	12.37
Theses	15.52	10.57	9.64	11.75
Webpages, websites	3.53	3.45	2.53	3.09
	Median age in years			
All materials	13	9	8	9
Book chapters	9	10	8	9
Media, maps, music	42	2	23	39
Monographs	16	11	10	12
Other	2	4	3	3
Proceedings, presentations	9	6	6	6
Serial articles	13	9	8	8
Theses	12	8	6	9
Webpages, websites	3	3	2	2
	Age of oldest			
All materials	406	218	216	406
Book chapters	84	218	119	218
Media, maps, music	201	12	23	201
Monographs	406	172	165	406
Other	75	115	82	115
Proceedings, presentations	67	83	67	83
Serial articles	116	166	216	216
Theses	55	46	54	55
Webpages, websites	12	14	13	14

or “webpages and websites.” Unique keywords negatively correlated with “book chapters” in STEM, “other” across all disciplines collectively and in STEM, and with theses in STEM. This means that if the use of book chapters, “other” sources, and theses were to increase (under the indicated discipline area conditions), the incidence of unique keywords could be expected to decrease. Conversely, unique keywords positively correlated with “proceedings and presentations” across all disciplines, meaning that if the use of proceedings and presentations were to increase, one could expect an increase in the appearances of unique keywords.

Keyword and LCSH variants, on the other hand, provided a number of statistically significant correlations. All material types collectively produced only one statistically significant correlation: in the social sciences, in which LCSH cross-references were variants of keywords. In this case, the implication is that as use of resources increases in the social sciences, one could expect to see a higher number of LCSH cross-references that were variant forms of keywords.

Book chapters resulted in the most correlations of statistical significance across all disciplines collectively, as well as separately in the arts and humanities and STEM, but not

Table 3. Correlations of Keyword and LCSH Variables with Material Type ($N = 38,317$)

Material Type	Discipline Area	Keyword and LCSH Variable	Correlation Coefficient	p Value	
All	Social Sciences	LCSH: cross-refs are variants of keywords	0.2	0.03	
Book chapters	All	Keywords: abbreviations of LCSH	-0.12	0.04	
		Keywords: all nonmatches	-0.17	0.00	
		Keywords: all variant types	-0.16	0.01	
		Keywords: variant spellings of LCSH	-0.15	0.01	
		LCSH: abbreviations of keywords	-0.13	0.03	
		LCSH: all nonmatches	-0.13	0.03	
		LCSH: all variant types	-0.17	0.00	
	Arts and Humanities	LCSH: variant spellings of keywords	-0.17	0.00	
		LCSH: all variant types	0.45	0.02	
		LCSH: variant spellings of keywords	0.45	0.02	
	STEM	Keywords: all nonmatches	Keywords: all nonmatches	-0.29	0.00
			Keywords: all variant types	-0.18	0.03
		Keywords: unique	Keywords: unique	-0.22	0.01
			LCSH: all nonmatches	-0.24	0.00
		LCSH: all variant types	LCSH: all variant types	-0.21	0.01
LCSH: variant spellings of keywords			-0.19	0.02	
Monographs	All	Keywords: all nonmatches	-0.13	0.03	
		LCSH: all nonmatches	-0.18	0.00	
		LCSH: all variant types	-0.14	0.02	
		LCSH: variant spellings of keywords	-0.16	0.01	
	Social Sciences	LCSH: cross-refs are variants of keywords	0.27	0.00	
	STEM	Keywords: variants of LCSH cross-ref	0.24	0.00	
		LCSH: all nonmatches	-0.21	0.01	
		LCSH: all variant types	-0.18	0.04	
Other	All	Keywords: all nonmatches	-0.14	0.02	
		Keywords: unique	-0.14	0.02	
	STEM	Keywords: all nonmatches	-0.2	0.01	
		Keywords: unique	-0.23	0.01	
		LCSH: all non-matches	-0.17	0.04	
Proceedings, presentations	All	Keywords: all nonmatches	0.13	0.03	
		Keywords: unique	0.12	0.04	
		LCSH: all non-matches	0.13	0.02	
		LCSH: all variant types	0.13	0.03	
Serial articles	All	Keywords: all variant types	0.13	0.03	
		LCSH: abbreviations of keywords	0.15	0.01	
		LCSH: all variant types	0.15	0.01	
		LCSH: variant spellings of keywords	0.12	0.05	
	Social Sciences	LCSH: abbreviations of keywords	0.2	0.03	
		LCSH: all nonmatches	0.21	0.03	
Theses	All	LCSH: all nonmatches	-0.13	0.03	
		Keywords: all nonmatches	-0.2	0.02	
	STEM	Keywords: unique	-0.2	0.02	
		Keywords: variants of LCSH cross-ref	0.27	0.00	

in the social sciences. Across all disciplines, book chapters negatively correlated under the variables that produced a statistically significant result: keyword and LCSH abbreviations of each other; keyword and LCSH variant spellings; all keyword variant types together, as well as all LCSH variant types together; and all keyword and LCSH nonmatches. In STEM, all the significant correlations were also negative (and slightly stronger than across all disciplines): unique keywords, all keyword variants together, and all keyword nonmatches, as well as LCSH variant spellings of keywords, all variant LCSH together, and all LCSH nonmatches. With book chapters across all disciplines, and in STEM, increased use of this resource type would be associated with a decrease in unique keywords and variant spellings of LCSH terms and a decrease in LCSH terms that were variants of keywords.

On the other hand, arts and humanities produced only two correlations, strongly positive and significant ones with LCSH variant spellings of keywords and all variant LCSH forms. These were the only statistically significant correlations found in the arts and humanities for the material-type-only set. In contrast to the other results for book chapters, these results indicate that an increased use of book chapters would be associated with a higher incidence of LCSH variants of keywords.

Monographs produced results across all disciplines in general, and the social sciences and STEM in particular, mostly with LCSH. Across all disciplines, variant LCSH spellings of keywords negatively correlated with monographs, as did all LCSH variant types together and all LCSH nonmatches. All LCSH variant types and all LCSH nonmatches also negatively correlated with monographs under STEM. In the social sciences, LCSH cross-references that were variants of keywords positively correlated with monographs. All keyword nonmatches correlated negatively with monographs across all disciplines collectively; however, keyword variants of LCSH cross-references correlated positively in STEM. For monographs, then, as with book chapters, an increased usage of these resources would be associated with a decrease in LCSH variants of keywords across all disciplines and in STEM. Unique and variant keywords would be expected to decrease. However, in the social sciences, an increase in the number of LCSH cross-reference variants of keywords would be indicated, and in STEM an increase in keyword variants of LCSH cross-references would be expected.

The "other" category of material type also resulted in negative correlations. Across all disciplines, as noted above, unique keywords correlated with this category, as did all keyword nonmatches. The same keyword variables produced slightly stronger results in STEM. STEM also produced a weak negative correlation between "other" and all LCSH nonmatches.

Proceedings and presentations, in addition to the positive correlation with unique keywords across all disciplines,

positively correlated with all keyword nonmatches and with all LCSH variant types and all LCSH nonmatches. These correlations are weak, but they imply that across all disciplines, an increase in the use of proceedings and presentations would be associated with an increase in variant forms of terms, both keyword and LCSH.

Serial articles produced no results in either the arts and humanities or STEM. Keyword variants taken together positively correlated with serial articles across all disciplines collectively. LCSH abbreviations and variant spellings of keywords, as well as all LCSH variant types taken together, also positively correlated with serial articles across all disciplines. The social sciences produced results between serial articles and LCSH abbreviations and all LCSH nonmatches. The implication in this case is that an increase in the use of serial articles may be associated with an increase in the use of keyword variants of LCSH terms and with LCSH variants, especially abbreviations, of keywords.

Theses as a material type produced perhaps the most mixed results of this set of correlations. Across all discipline areas, all LCSH nonmatches negatively correlated with theses. In STEM, unique keywords and all keyword nonmatches also negatively correlated with theses. However, keyword variants of LCSH cross-references positively correlated with theses. Thus, across all disciplines, an increase in the use of theses may be associated with a decrease in the incidence of unique and variant LCSH and unique and variant keywords of LCSH terms, but an increase in the appearance of keyword variants of LCSH cross-references also could be expected.

Of the forty-four statistically significant correlations found in this set, negative correlations considerably outnumbered the positive ones: twenty-eight negative (63.64 percent) to sixteen positive (36.36 percent). All of the positive correlations involved keyword or LCSH variants, mostly LCSH variants, and not unique keywords or LCSH separately. The positive correlations that fell in STEM (monographs and theses) all involved keyword variants of LCSH cross-references. That is, in STEM, increased usage of these two resource types is associated with more appearances of keywords that were variants of LCSH cross-references. Other positive correlations with keyword variants of some kind occurred in material types that appeared most often in STEM, that is, proceedings and presentations, and serial articles (which also contained positive correlations to LCSH variants of some type). All correlations that fell in the social sciences and in the arts and humanities were positive and involved some kind of LCSH variants. That is, in the social sciences and the arts and humanities, increased use of the correlated material type would suggest increased numbers of some kind of LCSH variant of keywords.

Examination of the results from the keyword- and LCSH-variable perspective revealed that nineteen of the

Table 4. Correlations of Keyword and LCSH Variables with Median Age of Each Material Type ($N = 38,087$)

Material Type (Median Age)	Discipline Area	Keyword, LCSH Variable	Correlation Coefficient	p Value
All	All	Keywords: unique	-0.14	0.02
		LCSH: all nonmatches	-0.17	0.00
		LCSH: all variant types	-0.13	0.03
		LCSH: variant spellings of keywords	-0.12	0.04
	Arts and Humanities	Keywords: abbreviations of LCSH	0.52	0.00
		Keywords: all variant types	0.43	0.02
	Social Sciences	Keywords: variants of LCSH cross-ref	0.2	0.03
		LCSH: cross-refs are variants of keywords	0.22	0.02
STEM	Keywords: unique	-0.17	0.04	
Media, maps, music	All	Keywords: all non-matches	-0.13	0.03
		Keywords: unique	-0.13	0.03
		LCSH: all nonmatches	-0.12	0.05
Monographs	All	Keywords: variants of LCSH cross-ref	0.14	0.02
	Arts and Humanities	Keywords: abbreviations of LCSH	0.58	0.00
		Keywords: all variant types	0.42	0.03
	Social Sciences	Keywords: variants of LCSH cross-ref	0.19	0.04
Other	All	Keywords: variants of LCSH cross-ref	0.22	0.00
		LCSH: cross-refs are variants of keywords	0.34	0.00
	Social Sciences	Keywords: variants of LCSH cross-ref	0.38	0.00
		LCSH: all variant types	0.23	0.01
		LCSH: cross-refs are variants of keywords	0.62	0.00
Proceedings, presentations	Social Sciences	LCSH: all non-matches	-0.2	0.03
Serial articles	All	LCSH: all nonmatches	-0.15	0.01
		LCSH: all variant types	-0.12	0.04
		LCSH: variant spellings of keywords	-0.12	0.04
	Social Sciences	LCSH: cross-refs are variants of keywords	0.19	0.04
Theses	All	Keywords: variants of LCSH cross-ref	0.21	0.00
	Social Sciences	Keywords: variants of LCSH cross-ref	0.18	0.05
	STEM	Keywords: variants of LCSH cross-ref	0.36	0.00
Webpages, websites	All	LCSH: all non-matches	-0.15	0.01
	STEM	LCSH: all non-matches	-0.19	0.02
		LCSH: all variant types	-0.19	0.03
		LCSH: variant spellings of keywords	-0.17	0.05

forty-four correlations (43.18 percent) involved unique keywords or keyword variants and twenty-five correlations (56.82 percent) involved LCSH variants of some kind. Within each of these variables, five of the keyword variables resulted in positive correlations while the remaining fourteen were negative. The LCSH variables broke down into eleven positive and fourteen negative correlations.

The discussion above that itemized the results by

material type contained descriptions of the negative correlations. General implications for the maintenance of LCSH will be summarized following the discussion of age of cited materials and unique keywords and LCSH.

Age of Cited Materials and Unique Keywords and LCSH

Table 4 summarizes the correlations of median age with

keyword and LCSH variable (unique keywords and LCSH, keywords and LCSH variants, all nonmatches). This table also collocates the results first by material type, then by discipline area, followed by keyword or LCSH variable and associated correlation coefficient and p value. In contrast to the results of material type only, the median age of book chapters did not produce any significant correlations. Median age across all materials produced correlations across all discipline areas collectively, as well as in each discipline area separately. However, as with the first set of correlations dealing with material type only, no significant correlations occurred between age and unique LCSH.

Median age across all material types and across all disciplines together produced negative correlations with unique keywords and with three LCSH variables (variant spellings of keywords, all variant types together, and all LCSH nonmatches). This means that as the use of older materials increases (that is, the median age goes up), the number of unique keywords as well as LCSH variants of keywords could be expected to decrease. A negative correlation between median age of all materials and unique keywords also occurred in the STEM discipline area.

However, the arts and humanities and the social sciences produced positive correlations with median age across all material types. In the arts and humanities, strong positive correlations occurred with keywords that were abbreviations of LCSH terms and with all keyword variant types taken together. In the social sciences, median age correlated with keyword variants of LCSH cross-references and with LCSH cross-references that were variants of keywords.

The median age of media, maps, and music produced negative correlations across all disciplines with unique keywords and all keyword nonmatches. A weak negative correlation also occurred with all LCSH nonmatches. An increase in the age of media, maps, and music is associated with a decrease in unique keywords and keyword variants of all types, as well as unique LCSH and LCSH variants of all types.

The median age of monographs resulted in positive correlations with keyword variables. Across all discipline areas collectively, and social sciences separately, the keyword variable was variants of LCSH cross-references. In the arts and humanities, the correlated variables were keyword abbreviations of LCSH and all keyword variant types. No significant correlations of median age of monographs occurred with either keyword or LCSH variables in STEM.

The median age of "other" material types correlated positively with the cross-reference variants of both keywords and LCSH in the social sciences and across all disciplines. In the social sciences, all LCSH variants of keywords also correlated positively with the median age of "other."

The median age of proceedings and presentations resulted in only one significant correlation. This occurred in

the social sciences with all LCSH nonmatches. This means that an increase in the median age of proceedings and presentations would be associated with a decrease in the number of unique LCSH and all LCSH variants of keywords for those resources.

Serial articles produced correlations between median age and LCSH variables. Across all disciplines, negative correlations occurred between LCSH variant spellings of keywords, all variant types together, and all LCSH nonmatches. The social sciences contained a positive correlation between median age of serial articles and LCSH cross-references that were variants of keywords. Across all disciplines in general, an increase in median age of serial articles would be associated with a decrease in unique LCSH and LCSH variants of keywords. However, in the social sciences, an increase in the median age of these resources would imply an increase in LCSH cross-references that were variants of keywords. No significant correlations of median age with keyword or LCSH variable occurred in the arts and humanities or STEM.

The median age of theses produced positive correlations between median age and keyword variants of LCSH cross-references across all disciplines collectively, as well as in the social sciences and STEM. In other words, older theses would be associated with an increase in the number of keyword variants of LCSH cross-references.

The median age of webpages and websites correlated negatively with LCSH variables. Across all disciplines together, the age of these resources negatively correlated with all LCSH nonmatches. In STEM, the correlations were with LCSH variant spellings of keywords, all LCSH variant types together, and all LCSH nonmatches. One would expect to associate older webpages and websites with fewer unique LCSH and LCSH variants.

Fewer statistically significant correlations occurred between median age of materials and the keyword and LCSH variables. The thirty-three correlations found under median age were evenly split: seventeen were positive (51.51 percent) and sixteen were negative (48.49 percent). As with the first set of correlations, all of the positive correlations were with some kind of keyword or LCSH variant, not with unique keywords or LCSH themselves. An even split between keyword and LCSH variables also occurred: sixteen of the correlations were with keyword variant of some type, of which twelve were positive and four were negative (including the three correlations with unique keywords only). The remaining seventeen correlations with LCSH variants broke down into five positive and twelve negative correlations. The negative correlations have been discussed above in the descriptions that were itemized by material type.

Correlations in the arts and humanities were all strongly positive and occurred with two keyword variables, that is, keyword abbreviations of LCSH and keyword variants of

all types together with the median age of all material types together and monographs. This means that as the age of all material types in general and monographs in particular increases, the number of keyword abbreviations of LCSH and the number of keyword variants of all types could be expected to go up as well. The remaining positive keyword correlations were all with keyword variants of cross-references: median age of all material types together in the social sciences; monographs in all disciplines together and in the social sciences; “other” in all disciplines together and in the social sciences; and theses in all disciplines together, in the social sciences, and in STEM. In these cases, as the age of these resources goes up, an increase in the number of keyword variants of LCSH cross-references could be expected.

The five remaining positive correlations all occurred in the social sciences. All but one were for LCSH cross-reference variants of keywords, which were with the median age of all material types together, monographs, “other,” and serial articles. For these resources, an increase in median age would be associated with an increase in the appearance of LCSH cross-reference variants. The outlier was a correlation of all LCSH variant types together with median age of “other” materials.

Implications for the Maintenance of LCSH

Implications for the maintenance of LCSH must be asserted with caution. This study was conducted with a limited number of documents from one institution, which became available in the same year, and generalizing the results to a larger population may not be possible. The findings are mixed and some are weak. However, two patterns may be identified.

Positive correlations in both sets of correlations can be associated with a need for more liberal use of cross-references. For example, where positive correlations occurred for the various material types, an increase in the use of a material type would be associated with a rise in the number of keyword variants of LCSH or LCSH variants (or LCSH cross-reference variants) of keywords. In all cases, particularly those of keyword variants, the incidence of these variants could be mitigated by an expanded presence of cross-references for established terms.

An exception to the pattern of positive correlations for material type occurred with proceedings and presentations, in which the findings indicated that more of this resource type could be associated with more unique keywords, more keyword variants, and more LCSH variants. In this case, more terms based on proceedings and presentations would need to be established. However, the principles of literary warrant may preclude this to some extent given the nature of this resource type, which tends to include first presentations of new research. New terms discovered in this resource type

may be judged as not sufficiently settled in the discipline to be proposed for addition to LCSH.

Positive correlations for median age of materials also may be interpreted as indicators for more cross-references. These are cases in which an increase in age was associated with an increase in the number of keyword or LCSH variants. The correlation set for median age had a consistent showing of positive correlations for keyword variants of LCSH cross-references, as well as LCSH cross-reference variants of keywords, which shows a need for a mechanism in LCSH to cover variant forms of established terms.

Negative correlations, which were the majority in the material-type correlation set and about half in the median-age correlation set, provide a different implication. If use of various materials increases or if the age of the materials goes up, the incidence of unique keywords, keyword variants, and LCSH variants, would be expected to decrease, implying more consistent or perhaps standardized use of terms. This would mean that terms, over time and as more resources appear, are more likely to be part of the canon of the discipline in which it appears and thus more likely to be established in LCSH.

Study Limitations

As noted above, this study was conducted with a limited number of documents (285) from one institution, which became available in the same year, and one cannot assume that the results would be the same with a larger population. Another potential problem is how subjects were sorted into the three broad categories (arts and humanities, social sciences, and STEM). The placement of some subjects that fell into the social sciences and STEM at Ohio State University might be handled differently at other institutions and by other researchers.

Areas for Additional research

More research is needed that examines the discipline areas in which successful LCSH proposals (for new terms and for changes or additions of cross-references) tend to be submitted and the materials being cataloged that inspire proposals. No studies have been published on the impact of the Cataloging-in-Publication program or PCC's Subject Authority Cooperative Program (SACO) on the maintenance of LCSH and the results of this study may suggest a direction for such research.

Conclusion

The goal of this study was to answer five research questions and test two assumptions by examining citations in

285 Ohio State University ETDs. The first question asked if broad disciplines evidence specific citation patterns. Data collected from citations found that broad disciplines (arts and humanities; social sciences; and STEM) have specific citation patterns. The arts and humanities made heavier use of monographs, STEM mostly used serial articles and rarely used monographs, and the social sciences fell between arts and humanities and STEM. These findings were consistent with those of earlier citation analysis research.

The second question sought to find identifiable patterns within disciplines when type and age of cited materials were examined. Cited materials were, on average, older in the arts and humanities group than in the social sciences or STEM groups, and the pattern held for median age as well. As with material type, these findings are consistent with earlier citation analyses.

An examination of correlations between type of cited materials and unique keywords or LCSH (the third question) found mixed results. No statistically significant correlations were found for unique LCSH with any material type, and none were found for media, maps, and music, or for webpages and websites. Unique keywords correlated positively with proceedings and presentations across all disciplines, but correlated negatively with “other” materials across all disciplines, and with book chapters, “other” materials, and theses in STEM. Other keyword and LCSH variants and nonmatches also were correlated with material types. The majority of these resulted in negative correlations, which implied that if usage of any given material type were to increase, the number of variant and nonmatched terms would go down.

The fourth question asked what correlations exist between age of cited material and unique keywords or LCSH. Again, no statistically significant correlations were found for unique LCSH. Book chapters also produced no significant correlations. Unique keywords correlated negatively with median age across all material types in all discipline areas collectively, in STEM, and with median age of media, maps, and music. As the age of these materials increases, the number of unique keywords may be expected to decrease. Keyword variants and nonmatches tended to correlate positively with median age while LCSH variants and nonmatches tended to correlate negatively. In other words, an increase in the age of materials cited may be associated with an increase in the number of keyword variants and nonmatches of LCSH, but the number of LCSH variants and nonmatches of keywords could be expected to decrease.

The expected differences between the arts and humanities and STEM did not fully materialize in the correlations. The extent of the contrasts between the discipline areas is mainly that more correlations with unique keywords, and keyword and LCSH variants, occurred in STEM than in

the arts and humanities. In the material-type correlations, the arts and humanities had positive correlations of LCSH variants with book chapters in opposition to the negative correlations of the same found in STEM. In the median-age correlations, the arts and humanities posted positive correlations of keyword variants with all material types in contrast to the negative correlation of unique keywords with all material types in STEM. Particularly notable is that neither the arts and humanities nor STEM contained statistically significant correlations of any variable with serial articles in either set of correlations.

The last question considered whether findings have implications for the maintenance of LCSH. The predominant implication is a need for more cross-references to established terms. To a lesser extent, a willingness to establish terms based on materials associated with new research, such as proceedings and presentations, as well as theses, also could reduce the rate of variance and nonmatch. Catalogers of ETDs may be well positioned to propose new LCSH because they are working with contemporary research in emerging fields. As these fields age and become established in the research canon, terms may be updated and expanded with cross-references to help future search and discovery.

Finally, the author tested two assumptions, both of which were confirmed. The arts and humanities do make heavier use of monographs, STEM makes very heavy use of serial articles and very little use of monographs, and the social sciences fall between arts and humanities and STEM. In addition, research in disciplines that cite monographs more often will show fewer unique keywords and LCSH.

This study contributes to the literature on the use of materials in graduate-level research and the patterns of use within disciplines, and it explored the extent to which currently available LCSH provide the level of specificity needed to describe and access these resources. More liberal use of cross-references is needed, as is a willingness to establish new terms early. The prevalence of unique keywords and keyword and LCSH variants and nonmatches indicates that LCSH is less effective at providing subject access to those material types that tend to present new research, such as proceedings and presentations, and theses, suggesting that LCSH and catalogers' proposals of new headings through literary warrant lag behind many current research topics.

References

1. C. Rockelle Strader, “Author-Assigned Keywords versus Library of Congress Subject Headings: Implications for the Cataloging of Electronic Theses and Dissertations,” *Library Resources & Technical Services* 53, no. 4 (Oct. 2009): 243–50.
2. *Ibid.*, 247.
3. National Information Standards Organization, *Guidelines for the Construction, Format, and Management of Monolingual*

- Controlled Vocabularies*, ANSI/NISO Z39.19-2005 (Bethesda, Md.: NISO, 2005), 16, webs.um.es/isgil/Z39-19-2005.pdf (accessed Feb. 3, 2012).
4. Ibid.
 5. Cataloger's Reference Shelf, LCSH—Principles of Structure and Policies for Application, 3.2 Literary Warrant, www.itsmarc.com/crs/mergedProjects/subjhead/subjhead/3_2_lcsh.htm (accessed May 9, 2012).
 6. Lois Mai Chan, *Library of Congress Subject Headings: Principles of Structure and Policies for Application* (Washington, D.C.: Library of Congress, 1990).
 7. Sevim McCutcheon, "Keyword vs Controlled Vocabulary Searching: The One with the Most Tools Wins," *The Indexer* 27, no. 2 (2009): 64.
 8. Library of Congress, Cataloging Policy and Support Office, "Library of Congress Subject Headings: Pre- vs. Post-Coordination and Related Issues," (Mar. 15, 2007; revised Dec. 20, 2007), 4, www.loc.gov/catdir/cpso/pre_vs_post.pdf (accessed Dec. 30, 2011).
 9. Library of Congress, Cataloging Policy and Support Office, "Making Subject Headings Proposals: H 187, When to Establish a New Topical Subject Heading," *Subject Headings Manual* (Washington, D.C.: Library of Congress, 2008): H187.
 10. Ibid.
 11. Anton J. Nederhof, "Bibliometric Monitoring of Research Performance in the Social Sciences and the Humanities: A Review," *Scientometrics* 66, no. 1 (2006): 81.
 12. Mu-hsuan Huang and Yu-wei Chang, "Characteristics of Research Output in Social Sciences and Humanities: From a Research Evaluation Perspective," *Journal of the American Society for Information Science & Technology* 59, no. 11 (2008): 1820–21.
 13. John Cullars, "Citation Characteristics of Monographs in the Fine Arts," *Library Quarterly* 62, no. 3 (1992): 325–42.
 14. Wolfgang Glänzel and Urs Schoepflin, "A Bibliometric Study of Reference Literature in the Sciences and Social Sciences," *Information Processing & Management* 35, no. 1 (Jan. 1999): 31–44.
 15. Anne L. Buchanan and Jean-Pierre V. M. Herubel, "Comparing Materials Used in Philosophy and Political Science Dissertations: A Technical Note," *Behavioral & Social Sciences Librarian* 12, no. 2 (1993): 63–70.
 16. Jeffrey D. Kushkowsky, Kathy A. Parsons, and William H. Wiese, "Master's and Doctoral Thesis Citations: Analysis and Trends of a Longitudinal Study," *portal: Libraries and the Academy* 3, no. 3 (2003): 459–79.
 17. Jeffrey Wehmeyer and Susan Wehmeyer, "The Comparative Importance of Books: Clinical Psychology in the Health Sciences Library," *Bulletin of the Medical Library Association* 87, no. 2 (1999): 187–91.
 18. Thomas W. Conkling et al., "Research Material Selection in the Pre-Web and Post-Web Environments: An Interdisciplinary Study of Bibliographic Citations in Doctoral Dissertations," *Journal of Academic Librarianship* 36, no. 1 (2010): 20–31.
 19. Margaret N. Rogers, "Are We on Equal Terms Yet? Subject Headings Concerning Women in LCSH, 1975–1991," *Library Resources & Technical Services* 37, no. 2 (1993): 181–96.
 20. Glänzel and Schoepflin, "A Bibliometric Study," 33.
 21. Nederhof, "Bibliometric Monitoring."
 22. Huang and Chang, "Characteristics of Research Output in Social Sciences and Humanities."
 23. Cullars, "Citation Characteristics of Monographs in the Fine Arts."
 24. Buchanan and Herubel, "Comparing Materials."
 25. Lowell L. Hargens, "Using the Literature: Reference Networks, Reference Contexts, and the Social Structure of Scholarship," *American Sociological Review* 65, no. 6 (Dec. 2000): 846–65.
 26. Kushkowsky, Parsons, and Wiese, "Master's and Doctoral Thesis Citations," 473.
 27. Conkling et al., "Research Material Selection."
 28. Strader, "Author-Assigned Keywords."
 29. Ibid., 247.
 30. André Du Laurens, *A Discourse of the Preservation of Sight: of Melancholike Disease, of Rheumes, and of Old Age*, trans. Richard Surphlet, 1599 (issued in facsimile by Oxford University Press in 1938).

Appendix. Academic Majors and Discipline Areas Represented by ETD Population Under Study

Discipline Area	College (at The Ohio State University, 2005)	Academic Unit (program, department, school, or college, as identified on title page of ETD)	Number of ETDs
Arts and Humanities	Arts	Art Education	2
		History of Art	1
		Music	9
		Theatre	2
	Humanities	French and Italian	2
		History	2
		Linguistics	4
		Philosophy	2
		Slavic and East European Languages and Literatures	2
		Spanish and Portuguese	2
Social Sciences	Business	Accounting and Management Information Systems	1
		Labor and Human Resources	1
	Education	Education	46
		Physical Activity and Educational Services	1
		Teaching and Learning	2
		Comprehensive Vocational Education	1
	Human Ecology	Human Development and Family Science	1
		Human Ecology	4
		Nutrition Graduate Program	1
	Social and Behavioral Sciences	Anthropology	2
		Communication	8
		Economics	5
		Geography	5
		Political Science	5
		Psychology	28
		Public Policy and Management	1
	Social Work	Sociology	2
Social Work		2	

STEM	Biological Sciences	Biophysics	2
		Evolution, Ecology, and Organismal Biology	1
		Molecular, Cellular, and Developmental Biology Interdisciplinary Graduate Program	5
	Engineering	Biomedical Engineering	4
		Chemical Engineering	10
		City and Regional Planning	2
		Civil Engineering	1
		Computer and Information Science	5
		Electrical Engineering	11
		Industrial and Systems Engineering	1
		Materials Science and Engineering	5
		Mechanical Engineering	7
		Welding Engineering	2
	Food, Agricultural, and Environmental Sciences	Agricultural, Environmental & Development Economics	3
		Animal Sciences	5
		Entomology	4
		Environmental Science	2
		Food Science and Nutrition	3
		Horticulture and Crop Science	3
		Human and Community Resource Development	2
		Natural Resources	3
		Plant Pathology	2
		Mathematical and Physical Sciences	Astronomy
	Biochemistry		2
	Chemical Physics		3
	Chemistry		9
	Geodetic Science and Surveying		1
	Mathematics		6
	Physics		9
	Statistics		7
	Medicine	Integrated Biomedical Science Graduate Program	2
		Molecular Virology, Immunology and Medical Genetics	1
		Neuroscience	1
Oral Biology		1	
Nursing	Pathology	1	
	Nursing	1	
Pharmacy	Pharmacy	1	
Public Health	Public Health	5	
Veterinary Medicine	Veterinary Clinical Sciences	1	
	Veterinary Preventive Medicine	5	
Total			285