International Federation of Library Associations and Institutions
1–15

DOI: 10.1177/0340035221989367
journals.sagepub.com/home/ifl

What we talk about when we talk about information literacy

Margaret S Zimmerman
School of Information, Florida State University, USA

Chaoqun Ni
University of Wisconsin–Madison, USA

Abstract

Information literacy skills are requisite to fulfilling one's potential and are highly connected to a good quality of life. However, the ways in which information literacy is discussed within the academic canon are largely unexplored, particularly as these conversations take place through different cultural lenses. The ways in which such cultures are grouped often rely on traditional methods of geographic clustering that are increasingly complicated by the disparate internal nature of societies. Using text analysis of a large bibliometric data set, this research is an attempt to examine how scholars around the world discuss information literacy in their publications. The authors pulled 3658 records with the exact term "information literacy" from the Scopus database. This data was analyzed for the most frequently employed words and phrases, and grouped by country. The authors then further grouped the countries by their levels of literacy, Human Development Index ranking, the average number of citations per article, and a metric created by the authors that assessed each country's progress in regard to the Sustainable Development Goals and population health. The results include a discussion of the differences in the ways that scholars from different cultures discuss information literacy, and a number of data visualizations to highlight differences in the data.

Keywords

Information literacy, information instruction, services, user populations, global perspectives, libraries, information, library and information science, bibliometrics, informetrics, webometrics, information systems, information retrieval, society, culture, development

Introduction

As demonstrated by research, it is clear that good information literacy skills are necessary in order to fulfil one's potential. Information literacy has been linked with maintaining good health, understanding literature related to current culture, and evaluating quality online information (Leung, 2010). It is also a fundamental component of civic participation and engagement (Lee, 2013). Poor information literacy has the capacity to negatively influence a person's career prospects, education, finances, and health (Lloyd et al., 2016), while positive information literacy is consistently linked to better health information acquisition and outcomes (Leung, 2010). Generally, levels of information literacy have been highly corroborated in scholarship across various cultures as having a relationship with quality of life (Avilés et al., 2016; Leung and Lee, 2012; Ukachi, 2015).

Despite its significance for factors regarding quality of life, there seems to be limited cross-cultural discussion on the manifestation of information literacy as a concept and practice in varied geographic contexts. In order to have a meaningful dialogue across international lines, the global scholarship on information literacy must be examined. This article attempts to do just that by using a bibliometric approach to shed light on how information literacy

Corresponding author:
Margaret S Zimmerman, School of Information, Florida State University, 142 Collegiate Loop, Tallahassee, FL 32306, USA.
Email: mzimmerman@fsu.edu

is regarded as a concept internationally. Ten years of Scopus database records—2010 through 2019—with the exact term "information literacy" in the title, abstract, or keywords were extracted and cleaned for a result of 3658 records. The data was grouped geographically, and a text analysis approach was utilized to extract the most common words and phrases from the titles and abstracts. The results were grouped with the most common keywords from each country. In an attempt to provide alternative ways to examine cultural contexts, the countries were then arranged according to various indicators selected by the authors. These included the Human Development Index (HDI), literacy level, the average number of citations per article produced in these findings, and a metric created by the authors that assessed each country's progress or change over approximately 20 years in regard to the Sustainable Development Goals (SDGs)—a statistic that is thoroughly described in the methods section.

The result of this approach is a number of data visualizations that display the most common terms to describe information literacy in different groupings of geographic regions, as well as a discussion of the themes present in each grouping. By seeing how researchers from different groupings of countries describe their scholarship on information literacy, it is the authors' hope that a deeper understanding of the term and its related practices can be developed. This work was also conducted during a cultural shift regarding how we think of, categorize, and interact with—and similarly resource and prioritize—differences in race, culture, and diversity within and across geographic lines. This work suggests, at an essential cultural moment, a reconsideration of the nature and conception of information literacy and understanding of geographic variation.

Background

Credit for coining the term "information literacy" is typically given to Paul Zurkowski (1974: 6) and stems from a line in a report he wrote to the National Commission on Libraries and Information Science in which he stated: "People trained in the application of information resources to their work can be called information literates." What is more significant in this report, however, is the way in which Zurkowski talks about information—not as a thing that exists as a separate entity but instead as an interaction of concepts and ideas with the mind of the user. In one regard, this is a step toward information-as-process instead of information-as-thing (Buckland, 1991). This also sets the standard for information literacy being an active pursuit, which is a progression of thought that has persisted. In a recent publication by Darin Freeburg (2017: 974), he states: "Conceptualizations of information literacy have shifted from a focus on identifying universal standards for finding information, to outlining dynamic skills, subjectivities, and creation processes that develop this information." It is this ability to pursue information-as-process through executing these dynamic skills that is the hallmark of individual success in an information-saturated society.

While the significance of information literacy ability has been made apparent through scholarship, the ways in which information literacy is discussed within the academic canon are not as clear. When scholars talk about information literacy, what do they talk about? Is it the same dialogue in different geographic clusters? There seems to be some thought within the scholarly community that information literacy and topics surrounding it are not globally synonymous. Virkus (2003) discusses the emergence of information literacy as a movement in the USA and, separately, Australia. She then draws comparisons with conceptions of information literacy in Europe. This is one of two articles uncovered in this research that examine information literacy specifically in a European context. The other looks only at the sociopolitical perspective and analyzes policy surrounding information literacy on the continent with the purpose of grouping by policy axes (Basili, 2011). Within Virkus's (2003: 3–4) article, there is mention of "information literacy developments in Canada, China, Japan, Mexico, Namibia, New Zealand, Singapore and South Africa." Another work tries to define information literacy specifically for the UK (Armstrong et al., 2005). Other articles attempt to dial in an official conception of what information literacy is without regard to geographic differences in the conception of the term (Owusu-Ansah, 2005; Špiranec and Banek Zorica, 2010). Yet another approach has been to acknowledge that there are scholars from different countries weighing in on the conceptualization of information literacy but then discuss their work as a homogenous group with no cultural distinctions (Rader, 2002). However, after exploring these examples, very little literature can be found that examines information literacy through the lens of different cultural contexts.

How would these cultural contexts be grouped? An additional issue that is highlighted in this research is how exactly to look at geographic clustering. Traditionally, scholars group country data by regions of the world or the income level of the society. Other metrics may be used for very specific reasons related

to a distinct line of inquiry. However, there is an increasing consensus of thought that such groupings of countries are unhelpful (Alonso et al., 2015). For example, a geographic region such as Western Europe or Southeast Asia can be ambiguous and difficult to define (Aguilera et al., 2007). This line of thought is upheld by Kyambalesa and Houngnikpo (2016) in their comprehensive discussion regarding the challenges of grouping African countries, as one example. To add complexity to such efforts, Fantom and Serajuddin (2016) discuss how World Bank income groups have changed significantly as the global economic landscape has altered, and that this classification system may be outdated. They contest that in individual countries there are many economies that may be disparate from each other, and so classifying a nation as middle-low income, for example, is painting with too wide a brush. Alonso et al. (2015) agree, stating that the internal economic situations of "developing" countries are much more diverse than when the original classification system was created; they list several other proposals for classification of countries such as per capita income, country indebtedness, state of governance, and the HDI, which is used in this study and assesses countries based on life expectancy, education, and per capita income.

The HDI was created to provide a metric that assesses people and their capabilities within a society. It examines the aptitude for a long and healthy life, including such metrics as life expectancy, education, and gross national income (United Nations Development Programme, 2020). It is a simplistic measure that provides one and is not inclusive of all aspects that are fundamental to a well-working society (Klugman et al., 2011). The HDI has come under fire for being too focused on economic growth at a weight that is disproportionate to the benefits of minor economic improvement. Likewise, the benefits of extra schooling are weighted disproportionately to the economic returns of that schooling (Ravallion, 2010). Despite these criticisms, the HDI is seen as a valuable assessment of geographic regions that constitutes a departure from the traditional means of measurement (Klugman et al., 2011).

Other metrics that were chosen for this work also quantify quality of life. The SDGs were adopted in 2015 by all United Nations member states and are intended to provide "a shared blueprint for peace and prosperity for people and the planet" (United Nations Department of Economic and Social Affairs, 2020). The goals evaluate measures relating to quality of life and the success of nations with regard to economic, health, civic, and environmental well-being. Each goal has a number of specific metrics that it assesses. For example, SDG1 is "End poverty in all its forms everywhere" and examines the number of people living on less than US$1.25 a day; the percentage of people of all ages living in poverty in all its dimensions according to national definitions; and the nationally offered social programs that provide aid. SDG2, "End hunger," looks at data the measures the number of stunted and wasted children who evidence acute undernourishment (United Nations Department of Economic and Social Affairs, 2020).

It is also worth mentioning that the SDGs were derived as successors to the Millennium Development Goals (MDGs), which were implemented by the United Nations in 2000 (World Health Organization, 2017). The SDGs are a continuation of and build on the progress attained by the MDGs. Looking holistically at the metrics of the SDGs, they examine nearly identical metrics to the MDGs, though ordered differently.

In conducting this study, the authors hope to answer the following questions: Are there differences in how scholars discuss information literacy across varied groupings of countries? Are there more compelling ways to group countries in order to have a deeper understanding of the academic dialogue happening internationally? And can a large corpus of academic writing give insight into how scholarship changes in relation to various indicators of well-being and scholarship in each grouping?

Methods

The following is a detailed presentation of the procedures of this study. Included in this is an explanation of the data retrieval from the Scopus database, the text mining process, and the application of ranking indices for the countries included in the research.

In order to retrieve as many records as possible, the Scopus database was selected because it proved to be the most comprehensive database available to the authors. At the time this research was being conducted, Scopus contained 24,600 serial publications and over 75 million records. Using the exact phrase search "information literacy" in titles, abstracts, and keyword fields, and limiting the result categories to journal articles and conference papers written in English (as the authors only speak English), the holdings of Scopus were searched for the years 2010 through 2019. The search produced 4364 results. Of these results, 706 were deleted because they did not have the author, abstract, index keywords, or geographic affiliations available. This still left a large sample of 3658 records. These records were then exported and combined into an Excel spreadsheet. The spreadsheet

included columns for the following data fields: title, abstract, year, keywords, citation counts, correspondence address, and institutional affiliation of the authors. After an initial cleaning of the data, the authors assigned a country of origin to each individual record based, first, on the correspondence address listed. If this particular piece of information was not available, the institutional affiliation was used instead. Within these robust results, it was found that scholarship from 93 countries was represented.

One of the obvious limitations of this method was that only articles in English were examined. This excluded 284 documents, including 136 in Spanish, 59 in Portuguese, 30 in Chinese, 30 in German, and 29 in other languages. Also, 95 countries were originally included in the data set. Two of these countries did not report to the World Bank and therefore had to be excluded from further analysis.

To understand the context of information-literacy-related research, this project utilized a text analysis approach to examine the titles and abstracts of the 3658 articles contributed by authors affiliated with 93 countries and territories. The necessary preprocessing and normalization of these texts was conducted prior to the text analysis, including the removal of stop words and lemmatization.

As a means to begin the process of cleaning the text, stop words were removed for the final analysis. Stop words are generally defined as being the most common words in a language (in our case, English). These words are removed before text processing because they are usually distracting and non-informative, and cost additional memory overhead during the text analysis process. There are multiple possible lists of stop words but, in this case, the stop words list of the Python Natural Language Toolkit, version 3.4, was utilized due to its popularity in the text analysis community.

Following the removal of stop words, with regard to further cleaning the text for proper analysis, lemmatization, another preprocessing step, was utilized. Lemmatization is the procedure of removing inflectional endings and returning to the base of a word, known as a "lemma" (Korenius et al., 2004). More specifically, this project employed the lemmatization procedure to group together the different inflected forms of a word so that they could be analyzed as a single term. It used the Python Natural Language Toolkit to transform all plurals to singular forms, as well as replace past-tense verbs with their present-tense counterparts. For example, "is," "are," "am," "were," and "was" were all transformed to their root verb "be" for analysis.

Once all of the noise was removed and the text was normalized, the n-gram technique was applied. The n-gram approach is a common technology for extracting key phrases based on their frequencies of occurrence in a bag of words. More specifically, an n-gram is a contiguous sequence of n items from a given sequence of text (Suen, 1979). Given a sentence, a list of n-grams can be constructed by finding pairs of words that occur next to each other (Dunning, 1994). This project used the n-gram approach in order to extract critical phrases from the text. This was based on the consideration that words co-occurring frequently are more likely to be a phrase of contextual meaning than those that co-occur infrequently. Additionally, the more frequently these phrases appear in the text, the more likely they are to play an important role. For this project, the text was explored with n ranging from 2 to 5. The Python Natural Language Toolkit was used to analyze the n-gram of the text.

After gathering all of the n-grams, the data was analyzed for relevancy. As a result, many of the n-grams proved to be irrelevant and were thus discarded. For example, the removed n-grams included "category have user," "develop new," "number have paper," and "consistently have data." These were all concordances that appeared with high frequency in the pool of data but were devoid of meaning for the purpose of this analysis. Both researchers examined the common concordances and agreed regarding which phrases were meaningful and which were considered noise. The phrase "information literacy" as it appeared on its own was also discarded because it was in every record as per the inclusion criteria. When "information literacy" appeared in concordance with other words, it was examined for relevance. The most common words and concordant phrases that were coherent, meaningful phrases were copied into an Excel spreadsheet and grouped by each of the 93 countries.

Keywords were also analyzed as part of this research, but were not subjected to the concordant phrase extraction described above, which was only used on titles and abstracts. Some keywords did happen to appear repeatedly as concordant phrases and were noted in the research. However, they were extracted from the data using different processes and were analyzed differently depending on how they were found in the data and therefore included by the authors of the scholarly works.

After identifying the meaningful concordant phrases, all the keywords were separated into individual cells in Excel and grouped by country. The keywords for each country were organized alphabetically into single rows to simplify the data analysis. The

results of these steps were 93 single rows in an Excel spreadsheet listing the country first, then all of the most common words and meaningful concordant phrases, followed by the keywords represented in the data set.

Once the data had been cleaned, ranking indices for the countries were determined and systematically developed. First, data was downloaded from the World Bank, UNESCO, and United Nations Development Programme's Human Development Reports websites. For each country represented in the Scopus data, data that represented specific indicators of the SDGs was extracted. The authors specifically chose the first five SDGs to address in this research. These are: "End poverty in all its forms everywhere" (SDG1); "End hunger, achieve food security and improved nutrition and promote sustainable agriculture" (SDG2); "Ensure healthy lives and promote well-being for all at all ages" (SDG3); "Ensure inclusive and equitable quality education and promote lifelong learning opportunities for all" (SDG4); and "Achieve gender equality and empower all women and girls" (SDG5) (United Nations Department of Economic and Social Affairs, 2019). In order to address current progress in relation to these goals, the following variables from 2000 and the most recent year's data available were extracted from the World Bank's data repository: the poverty headcount ratio (SDG1); undernourishment and the percentage of stunted children (SGD2); maternal mortality, under-five mortality, neonatal mortality, and the percentage of prevalence of HIV in the adult population (SGD3); the gross percentage of children enrolled in primary school (SGD4); and gender parity and fertility (SDG5) (World Bank, 2019a, 2009b, 2009c, 2009d, 2009e, 2009f, 2009g, 2009h, 2009i, 2009j). In some cases, the data from 2000 could not be isolated. When this occurred, the data from 1999 or 2001 was used instead. No data that was more than four years old was used for the most recent variables. Most of the data represented in this study was from 2017, as 2018 was not yet available for all of the variables. However, if it was available—for example, such as for the under-five mortality variable—it was used.

Admittedly, the data from the year 2000 predates the SDGs, which were created in 2012 and implemented in 2015. However, as stated in the background section, the SDGs were derived as successors to the MDGs, which were implemented by the United Nations in 2000 (World Health Organization, 2017). The first five SDGs were addressed by the first six MDGs, with similar metrics assessed as indicators. Therefore, the areas of progress that this research is attempting to examine have been prioritized by the United Nations for the entire period that the data represents. The authors chose to examine the first five SDGs instead of the MDGs simply because they represent the current terminology being used in this domain.

The percentage of improvement was calculated for each variable based on the difference between the earlier data and the later data. As an example of this, Romania had an under-five mortality rate of 21.9 per 1000 live births in 2000. This improved to 7.3 per 1000 live births in 2017 (World Bank, 2019d). This would be a 66.67% improvement. Based on the new variable of percentage of improvement over the approximate 16- to 17-year period, a rank order was applied to each country. The countries were ordered and given a number from 1 to 93 depending on where they fell in relationship to each other with regard to progress on each individual variable. The country that had the largest improvement in perinatal mortality would receive a 1; the country that had the lowest improvement would receive a 93. Once each country had a ranking for each variable, these rankings were averaged to give the country a specific ranking of overall progress. This ranking—henceforth referred to as the "progress" ranking—is unique to this research and will be used as another way to group the countries in the Scopus data set as showing high or low progress according to this metric.

The progress statistic created in this research was an experiment and is to be considered only as such. In designing this metric, which was an interesting and enjoyable endeavor, it must be remembered that countries that have the greatest disparities can also make the greatest gains, and therefore do not necessarily produce progress that is as progressive as a country—like Demark, for example—that had incredibly high quality-of-life metrics at the beginning of the evaluated time period. The first iteration of this work averaged the rate of change into a progress number. However, the results were heavily skewed by exactly the issue just mentioned—a country with a maternal mortality rate of 800 per 100,000 that decreased its rate by 30% would have a shockingly higher number than a country that went from 11 to 8 per 100,000. That is why rank order was assessed instead. While countries with larger initial disparities to adjust were still favored in the numbers, the stabilization of the figures into simple rankings made the differences less extreme. The purpose of the measure, of course, was to assess exactly this kind of progress, in which that decline of 30% should have a high ranking—without completely leveling out the success of a country with lower measures to begin with. This is described here in the methods in such detail because it

is a new attempt at methodology, in addition to the findings it creates.

Additionally, the other data that was collected in order to rank each country included their HDI (United Nations Development Programme, 2019), the literacy rate of the country (UNESCO Institute for Statistics, 2019), and the citation count from the data set in Scopus. In order to clarify this last variable, the citation count was averaged across the number of articles represented for each country. For example, Spain had 131 records represented in the data with 658 citations, which gave an average of 5.02 citations per article. The numeric variables, HDI, literacy level, and average citation counts were used to provide a rank order for each country.

The result of these steps was an Excel spreadsheet, which had simple numeric rankings for each country that could be aggregated into the highest and lowest ranked countries for each variable. Once the countries with the highest and lowest rankings for each variable were determined, the Scopus word data was extracted for each group. The extracted data was then ordered by the most frequently used words and phrases for each grouping. Once identified, the results of these groupings were compared between sets and examined for common themes, which are discussed in the results section. The USA was removed from the data set because the sheer number of keywords (16,898) skewed the results.

Finally, the most frequently found groups of words and the word counts were uploaded to Tableau 2019.3 in order to create descriptive and captivating data visualizations. These visualizations are also displayed in the results section of this article.

Findings and discussion

Table 1 shows each indicator, the countries with the highest and lowest values from the Scopus data set for that indicator, and the resultant most common terminology used in each data set with word counts. For the purposes of creating Table 1, duplicates across the indicators were removed, and the final lists represent unique values. Only the top-15 words are displayed for each grouping, though many more were examined for content depending on the size of the data set.

For the countries separated by HDI, the high HDI countries have the themes of research, methodology, and education. The words "assessment," "controlled study," "major clinical study," "questionnaire," and "surveys" all appeared frequently in this grouping. Additionally, terminology associated with education—such as "students," "education," "teaching," "e-learning," "curricula," "secondary schools," "blended learning," and "distance education"—was also found to be frequently occurring. Typical words associated with information science, such as "information seeking," "information dissemination," "information use," "knowledge management," and both "information sources" and "information systems," were also present and ranked high.

The most striking difference in the terminology represented from the countries with the lowest HDI rankings in the Scopus sample was the noticeably higher prioritization of words associated with health and geographic locations. For example, in the first 50 most common words from both samples, the countries with the highest HDI ratings listed three words associated with health and four countries. The countries with the lowest HDI rankings listed nine words associated with health and seven geographic locations. However, those numbers only demonstrate part of the significance of these findings. The words associated with health and geographic locations were much higher in frequency with respect to the sample for the lower HDI countries than they were for the higher HDI countries.

The terms representative of health for the lower HDI group included words such as "maternal mortality," "perinatal mortality," "pregnancy," "dystocia," and "uterine rupture," which are all terms associated with reproductive and maternal health. There was also the inclusion of words associated with chronic conditions such as "cholesterol," "albuminuria," and "hypertension."

Additionally, terms associated with education were also highly present in this grouping, although only the terms "education" and "library" appeared in the top 10. "Postgraduate students," "university libraries," "librarian," and "undergraduates" were all listed in the top-40 words and phrases from this data set. Figure 1 shows a tree graph displaying a comparison of the top words and phrases from each HDI grouping.

As is shown in Figure 1, the countries divided by highest and lowest literacy levels had many similarities in the data. This is exemplified by the data showing how both low and high literacy groupings had a strong emphasis on words associated with education and teaching.

However, the most frequently used words and phrases from the countries with the highest literacy levels had a more technical focus than the data from the countries with the lowest literacy levels. For example, the words "e-learning," "engineering education," "computer science," "information technology," and "microelectronics" all appeared in the top-40 words and phrases from the countries with high

Table 1. Listings of countries, most common unique terms, and counts per indicator.

HDI					
Top 10	Count	Most common language	Most common language	Count	Bottom 10
Norway	47	Higher education	Pakistan	29	Kenya
Switzerland	30	Australia	Nigeria	21	Zambia
Australia	28	Health literacy	Postgraduate students	9	Nepal
Ireland	26	Information seeking	University libraries	9	Pakistan
Germany	25	Questionnaire	Information literacy skills	7	Tanzania
Iceland	24	E-learning	Developing countries	6	Zimbabwe
Sweden	23	Assessment	Health	6	Nigeria
Singapore	22	Norway	Awareness	5	Uganda
The Netherlands	22	Priority journal	Ethiopia	5	Benin
Denmark	21	Curricula	Malaria	5	Ethiopia
	20	Social media	Medical	5	
	18	Controlled study	Risk factor	5	
	18	Surveys	Africa	4	
	16	Information dissemination	Albuminuria	4	
	16	Major clinical study	Attitude to health	4	
Literacy					
Top 10	**Count**	**Most common language**	**Most common language**	**Count**	**Bottom 10**
Latvia	50	Information science	Internet	12	Ghana
Cuba	31	Higher education	University libraries	11	Bangladesh
Estonia	17	Information management	Information literacy skills	9	India
Kazakhstan	17	Surveys	Postgraduate students	9	Morocco
Lithuania	16	Academic libraries	Information-seeking behavior	8	Nepal
Poland	15	Information behavior	Questionnaire	8	Zambia
Ukraine	14	Information culture	Risk factor	8	Nigeria
Russia	13	E-learning	Skill	8	Pakistan
Slovenia	12	Sustainable development	Ghana	7	Ethiopia
Croatia	11	Digital literacy	Electronic resources	6	Benin
	11	Information technology	Information technology	6	
	10	Computer science	Educational status	6	
	10	Curricula	Information resource	6	
	10	Information retrieval	Information retrieval	5	
	10	Information use	Information literacy skill	5	
Citations per article					
Top 10	**Count**	**Most common language**	**Most common language**	**Count**	**Bottom 10**
Bangladesh	26	Teaching	Knowledge management	4	Trinidad and Tobago
Israel	15	Academic libraries	Security	4	Morocco
Switzerland	15	Digital literacy	Access	3	Costa Rica
Australia	14	Information management	Competence	3	Jordan
New Zealand	13	Attitude to health	Data	3	Peru
Singapore	13	Digital game-based learning	Developing countries	3	Fiji
Hungary	11	Health knowledge	Competitiveness	2	Zimbabwe
Iceland	11	Motivation	Library	3	Ecuador
The Netherlands	10	Computer science	Decision making	2	Tunisia
Chile	10	Information use	Environmental scanning	2	Romania
	9	Access to information	Knowledge construction	2	
	9	Computers	Military education	2	
	9	Decision making	Public key infrastructure	2	
	9	Digital libraries	Social exclusion	2	
	9	Information retrieval	Spatial data	2	

(continued)

Table 1. (continued)

Progress Top 10	Count	Most common language	Most common language	Count	Bottom 10
Sri Lanka	26	Health literacy	Engineering education	26	France
Thailand	25	Psychology	Higher education	21	Israel
Benin	23	Information behavior	Curricula	20	Oman
Turkey	18	E-learning	Educational computing	18	Lithuania
Morocco	17	Academic libraries	Innovation	14	Lebanon
India	17	Information management	Information services	13	Kuwait
Ghana	17	Professional competence	University libraries	13	Germany
Nepal	17	Qualitative research	Information society	12	UK
China	17	Social media	Computer science	10	Cuba
Ethiopia	16	Curriculum	Information-seeking behavior	9	Bangladesh
	16	Information dissemination	Personnel training	8	
	16	Information retrieval	Teaching model	8	
	16	Media literacy	Information dissemination	7	
	15	Digital literacy	Library instruction	7	
	14	Information culture	Teacher	7	

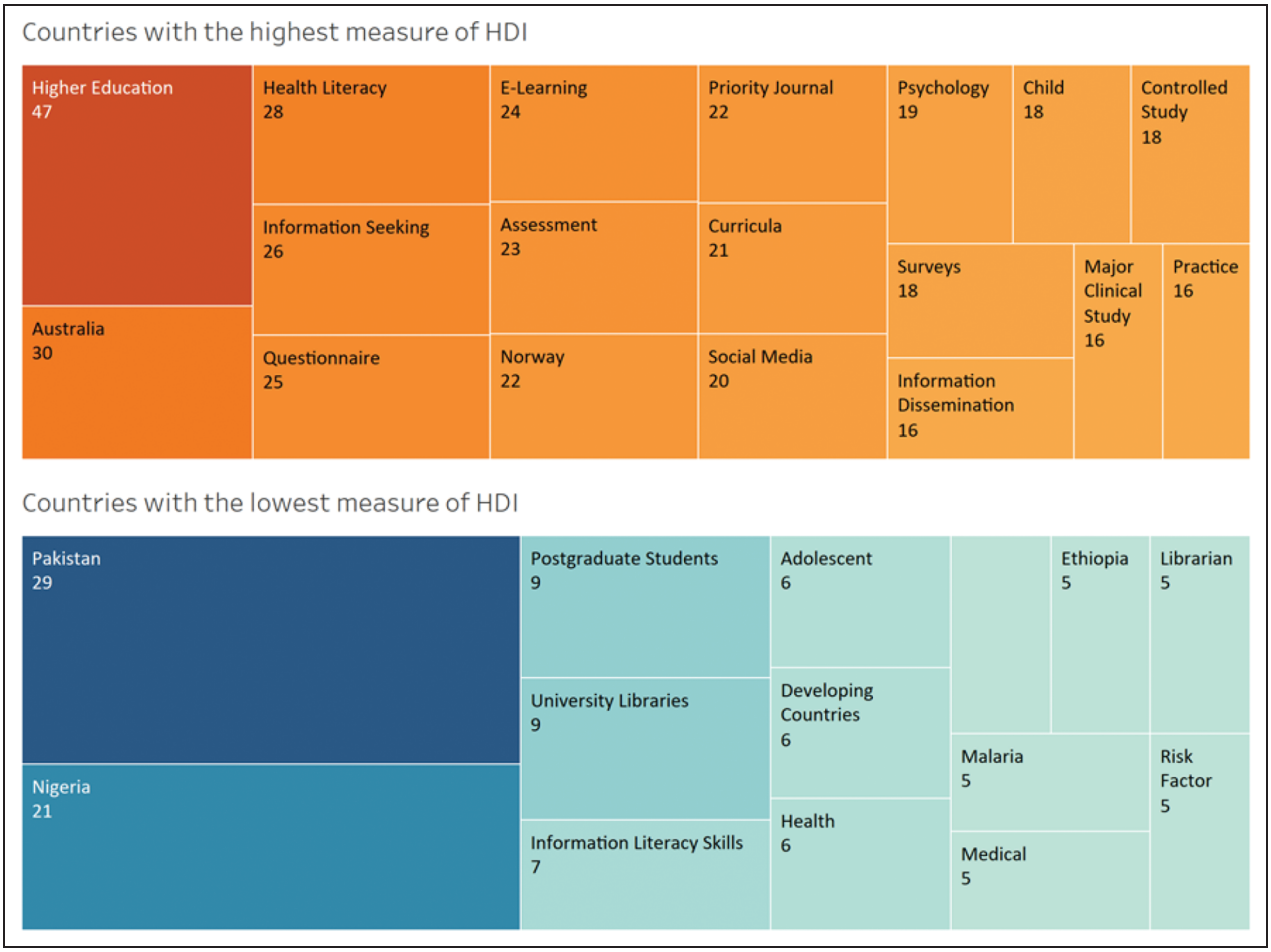

Figure 1. Comparison of the top words and phrases from countries with high and low HDI measures.

literacy levels. In fact, for the top-40 words and phrases for this group, 16 can be associated with technology or data science. "Digital divide," "information and communications technology use," "development," "international cooperation," and "sustainable development" were each ranked highly in the data for countries with higher literacy levels, suggesting a possible focus on literature exploring problems related to

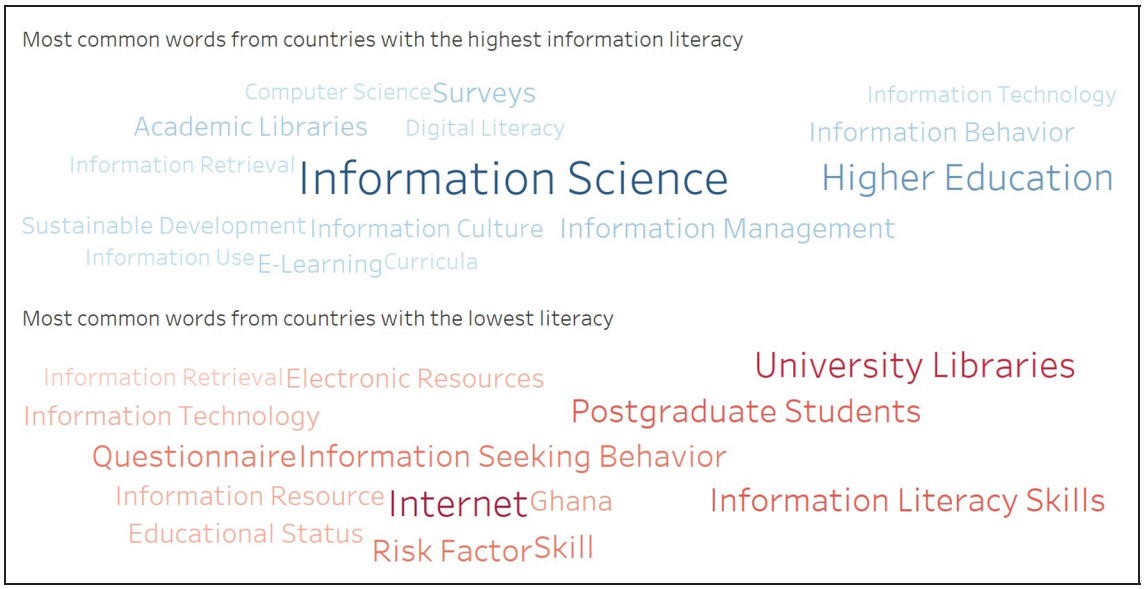

Figure 2. Wordles displaying the most frequent terms from countries with the highest and lowest literacy levels.

societies that traditionally suffer from digital-divide-related issues.

While the countries with lower literacy levels did have some technical terms represented in the data, the most frequently used terms were instead focused on education and research. The terms "postgraduate students," "undergraduates," "educational status," "teaching," and "higher education" were all among the top-30 words in the data, as well as terms such as "questionnaire," "risk factor," "major clinical study," and "cross-sectional study." The emphasis on words associated with information science and information seeking was more present in the word data from the countries with lower literacy levels than those with higher literacy levels. Unsurprisingly, "information-seeking behavior," "information science," "information retrieval," and "information literacy skill" were all among the top-30 terms. Words associated with health were represented in both data sets, but with much lower rankings than for HDI. Figure 2 represents a word cloud providing a visual image of some of the most common words from each grouping. The size of a word is directly related to how frequently the word was used. For example, in countries with the highest information literacy levels, the term "information science" was used more frequently than the term "information technology."

Ordering the countries by citations per article provided the most dramatic split of output between the number of articles that were included in each grouping. For example, the progress metric, which will be discussed next, calculated output from 10 countries that included 376 articles against output from 10 countries that included 317 articles. For literacy, the comparative groups were 163 articles against 191. For HDI, being the second largest difference, the total number of articles was 118 against 381. However, for the variable of average citations per article, the countries with the highest average citations included 267 articles versus 20 articles for the countries with the least number of average citations per article. This is in keeping with research which has found that the higher the scholarly output for a country, the more frequently the output from that country is cited (Pasterkamp et al., 2007).

That being said, the countries with the least number of citations per article still provided terms that were more frequently used. In this data were words that did not appear in other data sets—probably because this grouping provided an opportunity to highlight words and terms from some of the lesser-represented countries. Among these were terms such as "knowledge management," "security," "competitiveness," "military education," and "social inclusion," as well as "social exclusion." It is difficult to state a theme with data that appears twice among 20 articles because there simply is not a large enough corpus to assign consensus. What may be most of note related to this grouping is instead the incredible difference in scholarly output. This paucity seems to be

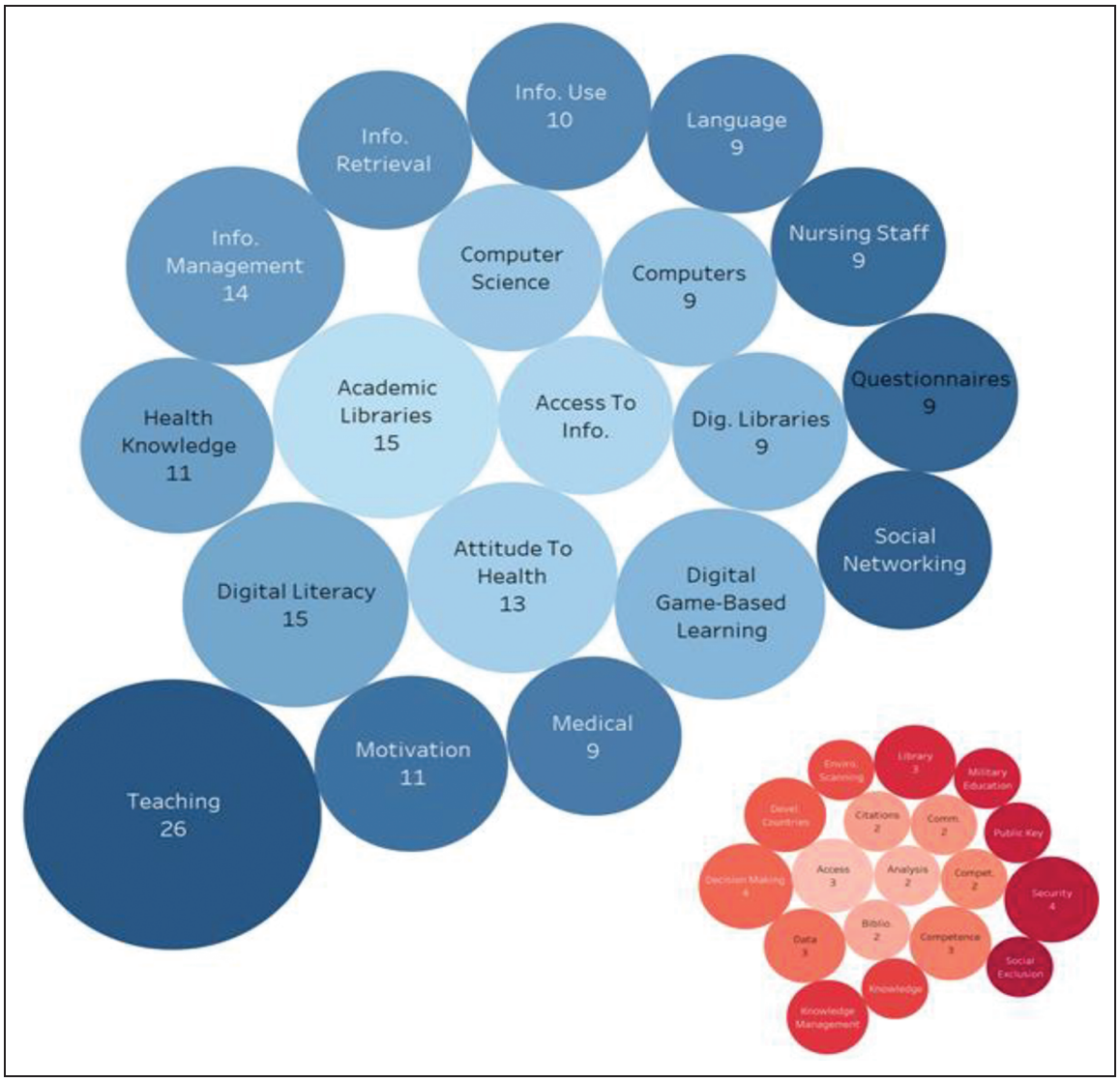

Figure 3. Word bubbles intended to show the scale of terms used between the citation groupings.

unrelated to other factors examined in this work. The countries examined in this sample fall across the spectrum of being classified by the World Bank as high, upper-middle, and lower-middle income. With the exception of Morocco, they fall around the median quartile for literacy level. Additionally, they are in the lower mid-range of the progress statistic and HDI, with the exception of Zimbabwe. Their one common factor is that none of the articles produced in this sample, including six articles from Romania, have been cited at all according to the Scopus database.

The grouping of countries with the most citations per article provided data that seemed to indicate foci on education and health. Some of the more prolific of the countries that appeared in this data set were also present in the data set for countries with the highest HDI, and because of this, many of the results matched. Like the countries with the highest HDI, words traditionally associated with information science were present and highly ranked, such as "information seeking," "information management," "information dissemination," "information use," and "access to information." Education-related terms were also found with similar frequencies, including "education," "teaching," "students," "academic libraries," "assessment," and "comprehension." However, unlike the grouping of data from countries with higher literacy, words associated with health were more frequent and more highly ranked. These included "attitude to health," "health knowledge," "medical," "nursing staff," "health education," and "medical education"—all of which appeared in the 40 most frequently listed terms.

In order to display the striking contrast between the word groupings from each data set, a circle chart is displayed in Figure 3. The smaller set is very difficult to read, but it is the contrast in the sizes of the two sets that is meaningful in this figure.

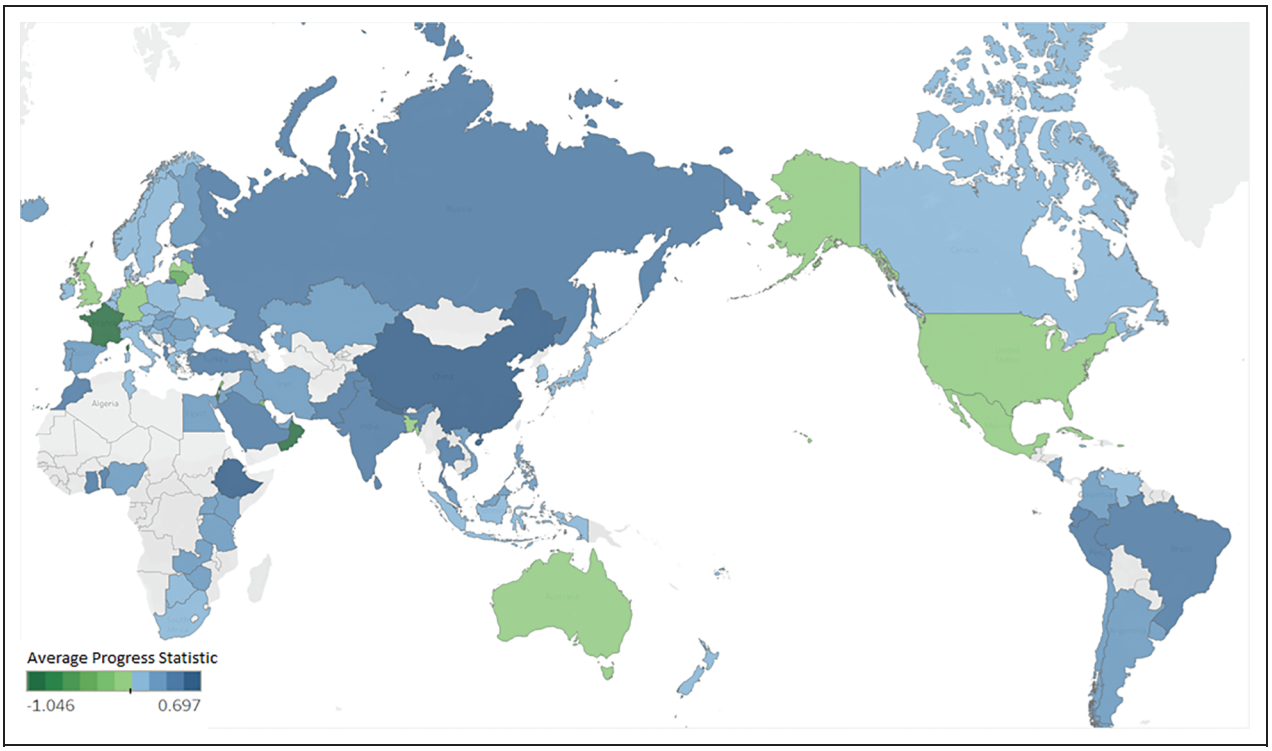

Figure 4. Word map displaying the progress statistic.

For the last comparative grouping, the countries were ranked according to the progress statistic described in the methodology section. As the results of this statistic accounted only for the percentage increase or decrease in specific metrics that were deemed indicative of the first five SDGs, the countries that are represented are strikingly dissimilar to the other ranking systems used in this research. For example, France, Germany, and the UK actually had negative progress statistics due to declines in areas such as gender parity, school enrollment, and neonatal mortality. The USA, before it was excluded from the data, also had a negative progress statistic.

However, countries with traditionally lower indicators of progress, such as HDI, have had high levels of reduction in some of these areas—especially if their numbers were low to start with. Using Ethiopia as a case study, it has reduced neonatal mortality by 42%, maternal mortality and under-five mortality by 61% each, and HIV infections by 58%, and increased gender parity by 88%. These significant gains, along with other improvements across the board, gave it the highest ranking of progress according to SDG indicators over the last 17 years, despite the fact that many of Ethiopia's current mortality rates are significantly higher than the high-income countries that ranked toward the bottom of this metric.

To be clear, the progress metric created in this research is not meant to capture the most successful countries in terms of the SDGs. There are many ways to rank the world's healthiest and least healthy societies, and they often produce similar lists. Instead, this ranking system is intended to display countries showing demonstrable effort in improving significant indicators of quality of life—even if those countries still have years of progress to work toward. The world map in Figure 4 has each country represented in this data set shaded in. Countries that are blue have a positive progress score. Countries that are green have a negative ranking. The darkness of each country is reflective of the distance from zero of their ranking.

With this in mind, the purpose of this research is not to assign labels to groups of countries, but instead to see how scholars from these groups discuss information literacy in their work.

With each variable, one grouping of countries had a much higher focus on education than the others. With this particular grouping, the countries with the lowest progress statistics listed terms associated with education for almost every one of the 20 most highly ranked terms. The exceptions to words associated with education are still arguably in the same vein, as "innovation," "information services," and "information dissemination" can absolutely be linked to education, and particularly to higher education. As the most frequently used word list expanded past 20, terms synonymous with research were present. "Cross-sectional study," "citation analysis," "factor

Figure 5. A comparison of terms used by scholars from higher and lower progress countries.

analysis," and "reliability" were among the frequently used terms.

The high-progress countries were less thematic. The resultant list of highly ranked words in this grouping centered around health, information science, and research. It was somewhat surprising that in a corpus of literature discussing information literacy, "health literacy" was the most frequently used term, followed by "psychology," which had not previously been present in the highly ranked data. In fact, for this grouping, the outliers that had not been resultant in the other analysis may be the most significant results. While the other group lists frequently repeated data, this grouping produced the most unique results. Some of these were "professional competence," "qualitative research," "consumer health information," "collaboration," "critical thinking," and "evidence-based practice."

Regardless of the words that were unique to this particular list or those that showed frequent crossover with the other lists, the data as a whole was most typically representative of some aspect of health, education, or research. This was interesting as these are the countries that have demonstrated the most progress with health- and education-related indicators of well-being in the data set. Figure 5 provides a visual display of the data for these groupings.

Conclusion

In examining international scholarship on information literacy, it is important to understand how scholars from different cultural backgrounds describe their academic production. The goal of this research was to find different ways to synthesize academic literature related to information literacy and, in doing so, determine alternative ways to look at how countries can be grouped in respect to a large data set organized geographically. It was the hope of the authors that by analyzing a large corpus of literature spanning 10

years of scholarly production, themes would emerge in the words, phrases, and keywords that were most frequently represented across different geographic spaces. It was also the goal of this project to forgo traditional country groupings, such as regions of the world or country income level, and find alternative ways of examining the data according to various societal indicators. This research is well in line with current trends in information science and librarianship which seek to promote international cooperation and understanding in an effort to find collaborative responses to global problems (Lor, 2019). The authors have conducted this research with these goals in mind in their belief that reaching understanding between cultures is the first step in working toward collaborative progress, and in agreement with the belief that the mission of information professionals is to improve society by facilitating knowledge creation (Lankes, 2016).

These goals have been met with varied success. There were few striking dissimilarities. The themes of education, health, technology, and research were present in the majority of the groupings, with the differences lying in the concentration of these themes between the data sets. Some unique themes did emerge, such as the emphasis on specific chronic and reproductive health issues in the grouping of countries with a lower HDI. Additionally, the unique terminology of the countries with the least average citation counts was of interest, as it pertained to considerations present in low-income countries such as access and infrastructure. While the data set for this grouping was not large enough to draw strong conclusions, the results that are present indicate that further examination may be merited.

It was also of interest to note that the countries with the lowest literacy rates, the lowest progress statistics, and the lowest HDI all had a high focus on education. The grouping of high HDI countries and the countries with the highest average citation counts also had highly repetitive terms associated with education. While, perhaps unsurprisingly, education was prioritized to some degree by nearly every data set, there was an uneven split as to whether it was more frequently repeated within data sets that represented higher or lower indicators of quality of life.

This phenomenon was also true, to a lesser extent, with health terms. Health-related language was the most highly ranked of the lowest HDI countries and was present, though poorly represented, in the data set produced by the grouping of countries with the lowest literacy levels. However, health flipped to be highly present in the data sets produced by the countries with the highest progress statistics and the highest average citation counts.

Finally, the method of grouping countries by different indicators provided an alternative way to look at the terminology used by scholars, as did the progress statistic introduced in this research. It is often traditional to examine scholarly output by regional categories or terms like "developing" or "low-income" versus "developed" or "high-income" countries. Instead, this research examined literature based on different indicators of quality of life and societal health in an effort to see, once arranged in such a manner, if the work produced by academics in these groupings can provide insight into what scholars in different societies are discussing when they talk about information literacy. In segregating the literature in these different ways, it was demonstrated that there are in fact differences in scholarship produced by countries with high HDI versus low HDI, high literacy versus low literacy, and those that have made marked progress with the SDGs versus those that have made no progress or whose progress has even declined. The metric of citations per article did not produce enough data to conclusively designate themes. However, this does align with previous research on the relationship between scholarly output and citation counts.

Limitations

Admittedly, there are substantial limitations to this work. One is the data sets that were used. For example, 21 countries do not report their levels of literacy and could not be analyzed for that particular grouping. The World Bank data, while the best available to the authors, also has its own set of limitations regarding quality. Some of the data sets that were downloaded were incomplete, and because some data is self-reported by each country, occasionally some numbers appeared to be inflated and unlikely to be accurate given other public information about the well-being of the citizens of that particular nation. In these circumstances, the authors deferred to the expertise of the World Bank and used the numbers provided. Additionally, the authors only looked at scholarship written in English, which omits any research published on this topic in other languages. This was necessary owing to the limitations of the authors' ability, but was still limiting. It is also worth noting that articles that are written in good English are more likely to be accepted for publication. Worthwhile research that is not written clearly and with grammatical accuracy is more likely to be rejected.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The authors received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Margaret S Zimmerman https://orcid.org/0000-0002-0518-1707

References

Aguilera RV, Flores RG and Vaaler P (2007) Is it all a matter of grouping? Examining the regional effect in global strategy research. In: Tallman SB (ed.) *A New Generation in International Strategic Management*. Cheltenham, UK: Edward Elgar, 209–228.

Alonso JA, Cortez AL and Klasen S (2015) LDC and other country groupings: How useful are current approaches to classify countries in a more heterogeneous developing world? In: Alonso JA and Ocampo JA (eds) *Global Governance and Rules for the Post-2015 Era: Addressing Emerging Issues in the Global Environment*. London: Bloomsbury Academic, 347–388.

Armstrong C, Boden D, Town S, et al. (2005) Defining information literacy for the UK. *Library and Information Update* 4(1–2): 22–25.

Avilés JM, Larghi SB and Aguayo MAM (2016) The informational life of the poor: A study of digital access in three Mexican towns. *Telecommunications Policy* 40(7): 661–672.

Basili C (2011) A framework for analyzing and comparing information literacy policies in European countries. *Library Trends* 60(2): 395–418.

Buckland MK (1991) Information as thing. *Journal of the American Society for Information Science* 42(5): 351–360.

Dunning T (1994) Statistical identification of language. Technical Report MCCS 940-273, Computing Research Laboratory, New Mexico State University, USA.

Fantom N and Serajuddin U (2016) The World Bank's classification of countries by income. Policy Research Working Paper 7528, The World Bank, USA, January.

Freeburg D (2017) A knowledge lens for information literacy: Conceptual framework and case study. *Journal of Documentation* 73(5): 974–991.

Klugman J, Rodríguez F and Choi HJ (2011) The HDI 2010: New controversies, old critiques. *Journal of Economic Inequality* 9(2): 249–288.

Korenius T, Laurikkala J, Järvelin K, et al. (2004) Stemming and lemmatization in the clustering of Finnish text documents. In: *Proceedings of the 13th ACM international conference on information and knowledge management*, Washington DC, USA, November 2004, pp. 625–633. New York, NY, United States.

Kyambalesa H and Houngnikpo MC (2016) Regional economic groupings. In: *Economic Integration and Development in Africa*.: Routledge, 107–126.

Lankes RD (2016) *The New Librarianship Field Guide*. Cambridge, MA: MIT Press.

Lee A (2013) Literacy and competencies required to participate in knowledge societies. In: Lee A, Lau J, Carbo T and Gendina N (eds) *Conceptual Relationship of Information Literacy and Media Literacy in Knowledge Societies*. Paris, France: United Nations Educational, Scientific and Cultural Organization (UNESCO), 3–62.

Leung L (2010) Effects of Internet connectedness and information literacy on quality of life. *Social Indicators Research* 98(2): 273–290.

Leung L and Lee PS (2012) Impact of Internet literacy, Internet addiction symptoms, and Internet activities on academic performance. *Social Science Computer Review* 30(4): 403–418.

Lloyd A, Lipu S and Kennan MA (2016) On becoming citizens: Examining social inclusion from an information perspective. *Australian Academic and Research Libraries* 47(4): 304–315.

Lor PJ (2019) *International and Comparative Librarianship: Concepts and Methods for Global Studies*. Berlin/Boston: De Gruyter.

Owusu-Ansah EK (2005) Debating definitions of information literacy: Enough is enough! *Library Review* 54(6): 366–374.

Pasterkamp G, Rotmans J, De Kleijn D, et al. (2007) Citation frequency: A biased measure of research impact significantly influenced by the geographical origin of research articles. *Scientometrics* 70(1): 153–165.

Rader HB (2002) Information literacy 1973–2002: A selected literature review. *Library Trends* 51(2): 242–259.

Ravallion M (2010) Troubling tradeoffs in the Human Development Index. Policy Research Working Paper 5484, The World Bank, USA, November.

Špiranec S and Banek Zorica M (2010) Information literacy 2.0: Hype or discourse refinement? *Journal of Documentation* 66(1): 140–153.

Suen CY (1979) N-gram statistics for natural language understanding and text processing. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 1(2): 164–172.

Ukachi NB (2015) Exploration of information literacy skills status and impacts on the quality of life of artisans in Lagos, Nigeria. *New Library World* 116(9–10): 578–587.

UNESCO Institute for Statistics (2019) Education: Literacy rate. Available at: http://data.uis.unesco.org/ (accessed 1 October 2019).

United Nations Department of Economic and Social Affairs (2019) Sustainable Development Goals. Available at: https://sustainabledevelopment.un.org/ (accessed 31 October 2019).

United Nations Department of Economic and Social Affairs (2020) The 17 goals. Available at: https://sdgs.un.org/goals (accessed 17 December 2020).

United Nations Development Programme (2019) Human development data. Available at: http://hdr.undp.org/en/data (accessed 1 October 2019).

United Nations Development Programme (2020) Human Development Index. Available at: http://hdr.undp.org/en/content/human-development-index-hdi (accessed 17 December 2020).

Virkus S (2003) Information literacy in Europe: A literature review. *Information Research* 8(4). Paper 159. Available at: http://informationr.net/ir/8-4/paper159.html

World Bank (2019a) Fertility rate, total (births per woman). Available at: https://data.worldbank.org/indicator/SP.DYN.TFRT.IN?view=chart (accessed 1 October 2019).

World Bank (2019b) Maternal mortality ratio (modeled estimate, per 100,000 live births). Available at: https://data.worldbank.org/indicator/SH.STA.MMRT?viewchart (accessed 1 October 2019).

World Bank (2019c) Mortality rate, neonatal (per 1,000 live births). Available at: https://data.worldbank.org/indicator/SH.DYN.NMRT?view=chart (accessed 1 October 2019).

World Bank (2019d) Mortality rate, under-5 (per 1,000 live births). Available at: https://data.worldbank.org/indicator/SH.DYN.MORT?view=chart (accessed 1 October 2019).

World Bank (2019e) Poverty headcount ratio at national poverty lines (% of population). Available at: https://data.worldbank.org/indicator/SI.POV.NAHC?viewchart (accessed 1 October 2019).

World Bank (2019f) Prevalence of HIV, total (% of population ages 15–49). Available at: https://data.worldbank.org/indicator/SH.DYN.AIDS.ZS?view=chart (accessed 1 October 2019).

World Bank (2019g) Prevalence of stunting, height for age (% of children under 5). Available at: https://data.worldbank.org/indicator/SH.STA.STNT.ZS?view=chart (accessed 1 October 2019).

World Bank (2019h) Prevalence of undernourishment (% of population). Available at: https://data.worldbank.org/indicator/SN.ITK.DEFC.ZS?view=chart (accessed 1 October 2019).

World Bank (2019i) School enrollment, primary (% gross). Available at: https://data.worldbank.org/indicator/SE.PRM.ENRR?view=chart (accessed 1 October 2019).

World Bank (2019j) School enrollment, secondary (gross), gender parity index (GPI). Available at: https://data.worldbank.org/indicator/SE.ENR.SECO.FM.ZS?viewchart (accessed 1 October 2019).

World Health Organization (2017) Millennium Development Goals (MDGs). Available at: https://www.who.int/topics/millennium_development_goals/en/ (accessed 31 October 2019).

Zurkowski PG (1974) *The information Service: Environment Relationships and Priorities*. Washington, DC: National Commission on Librarians and Information Science.

Author biographies

Margaret Zimmerman is an assistant professor in the School of Information at Florida State University. For her education, completed a doctorate in Library and Information Science from the University of South Carolina. She also has a Master's in Information Systems and an MLIS, both from Drexel University. Her research areas of interest focus on the health-seeking behaviors and patterns of women of disadvantaged populations. Professor Zimmerman studies the impact that information access, information literacy, and reading and literacy has had in affecting the health and well-being of the women she studies. She works with large datasets and at a smaller, more personal scale. Special interest is taken in maternal health outcomes. Zimmerman's recent research includes a comparative study between assessments of information literacy, health literacy, and the library science methodology Information Horizons Mapping, a large-scale bibliometric analysis of themes in global scholarship.

Dr. Chaoqun Ni is an Assistant Professor at the Information School, University of Wisconsin-Madison. Before that, she worked at Simmons University and the University of Iowa. Dr. Chaoqun Ni got her PhD in Information Science with a minor in Statistics from Indiana University Bloomington. For research, she studies science, scholarship, and the scientific workforce using massive data to inform decision-making on science policies. She mostly teaches in the areas of information technology, database, data visualization, and research methods.