Linguistic Diversity Index: A Scientometric Measure to Enhance the Relevance of Small and Minority Group Languages

Václav Linkov1, Kieran O’Doherty2, Eunsoo Choi3, and Gyuseog Han4

Abstract

Current scientometric indexes do not encourage the linguistic diversity of sources cited in academic texts and researchers are not motivated to cite texts written in smaller languages. This diminishes the cultural diversity of the sources cited and limits the representation of small and indigenous cultures. This text proposes a scientometric measure designed to encourage the linguistic diversity of sources cited in articles, books, and papers. The Linguistic Diversity Index is based on two stipulations: (a) the more linguistically diverse the sources, the higher the score, and (b) the rarer the languages cited, the higher the score. If such a metric were used for the evaluation of social science and humanities journals, it would encourage the publication of papers that cite ideas from rarely represented cultural groups such as indigenous nations, ethnic groups from small countries, and other linguistic groups that have been omitted from mainstream scientific discourse. This might help to produce new research, which would help to improve the situation for these groups and create an epistemology that is more just to small cultural groups.

Keywords

scientometrics, bibliometrics, linguistic diversity, cultural diversity, research evaluation

Research published in international social science journals is not culturally diverse. The vast majority of published studies come from affluent and developed countries, and small cultures and the cultures of indigenous people are neglected (Arnett, 2008). This might be because the cultural diversity of published research is not rewarded by the people who evaluate research. Scientometric indicators currently used in the evaluation of scientific research blind policy setters to many issues (Barré, 2019), of which cultural diversity is one. Current indicators do not reward cultural diversity, and therefore, this diversity does not exist as an important issue for people who use these indicators as evidence of research quality. This article aims to help diminish this problem by offering a scientific indicator that measures cultural diversity in academic texts.

Scientific output is variable, so it is not easy to measure (Pedersen et al., 2020). This is especially true when it comes to the measured output of the social sciences and humanities (Reale et al., 2018) because they are heterogenous (Fanelli & Glänzel, 2013). To measure this output, it is necessary to first evaluate “what is and is not desirable change in academic knowledge” (Hammersley, 2014, p. 351). In the social sciences and humanities, research concerns its influence on society (Ochsner et al., 2013), so this desirable change might be to “produce improvements in relation to the stated goals of society” (Reale et al., 2018, p. 300). Such goals might be to improve the cultural capital of society (Donovan, 2008) and cultural diversity. Nevertheless, research in the social sciences is mainly evaluated by the same standards as research in the natural sciences: by the number of citations the research receives (Hoblík, 2020), which does not support cultural diversity.

The academic world is more dependent on English-language sources than it was 30 years ago. Scholars are forced to become part of the English-language publication system, which values themes relevant for the dominant cultures of the English-speaking world (Dahler-Larsen, 2018).
If scholars focus on themes that are specific for local cultures and of low interest for this larger academic world, they might be disadvantaged compared with those who work on themes that are highly cited in the English-speaking world. When scientific output is judged by the number of citations in citation databases, publications in languages other than English are discriminated against (Towpik, 2015). Articles written in English that are published in multilanguage journals are cited more often than articles written in other languages and published in the same journal (Diekhoff et al., 2013). Scientists want to increase their number of citations (Furnham, 2020), and this forces them to publish in English and not in their native language (López-Navarro et al., 2015). Journals have a similar motivation (Chavarro et al., 2018), so they follow the same pattern, abandoning local languages and publishing in English (Dinkel et al., 2004). For an increased citation rate, it is also beneficial to get co-authors from developed countries (Meneghini et al., 2008) and publish in journals in these countries (Strehl et al., 2016). The advantage for those who want to increase their citation count is to cite texts written in English. Gong et al. (2019) have shown that Chinese scholars’ publications that cite more English-language sources are more cited than publications that cite more sources in Chinese. However, abandoning publishing in a native language and abandoning the citation of sources in non-English languages contribute to the impoverishment of native cultures (Filippov, 2016) and a decrease in cultural diversity. This can leave indigenous and small cultures less researched.

When diversity is mentioned in the context of scientometrics, it is usually diversity in disciplinary approaches (Zitt & Bassecoulard, 2008) and not cultural or linguistic diversity. This is because the prevalent methods of scientific evaluation were developed with a focus on “narrower” disciplines. Scientific disciplines differ according to their narrow focus (e.g., details and mechanisms) or wide focus (e.g., systems and larger processes). For example, cognitive psychology and experimental psychology are narrower; social psychology is broader. Some researchers in narrower fields think that the broader fields lack a solid basis upon which to build theories, whereas some of those in broader fields think that the narrow fields lack questions of real value and lose sight of the forest for the trees (Peterson, 2017). Research in the social sciences and humanities benefits from understanding specific cultures (Choi & Han, 2000), but researchers from narrow fields might consider the diversity of cultural experience to be less important than those from broad fields. Problems with the current indicators that were developed with such a narrow focus lead to the necessity of developing new and better indicators (Barré, 2019). Therefore, it might be beneficial if broader disciplines would use newly developed scientometric indexes that are different from those used in the narrower disciplines, so that these new indices would enhance the value of cultural diversity. This type of metric is proposed in this text.

### Linguistic Diversity Index (LDI)

Research in the social sciences and humanities often studies phenomena that vary among cultures. If a researcher wants to study some culturally specific phenomenon in a particular cultural group, they need to improve their understanding of that culture (Linkov, 2014). This is best fulfilled if the researcher knows the opinions of the members of this cultural group and includes these members in the research. Indeed, the more widespread inclusion of the opinions of small cultural groups in the international research arena would allow them to influence the usage of the research questions and the methods created for larger groups, which are not necessarily fair in the treatment and assessment of these small groups (Urbánek & Čeněk, 2019). A good understanding of various cultural groups in published research should, therefore, be supported.

The managers of research institutions have to rely on the quantitative indicators of scientific output because they are not professionals in the specific fields (Kuleshova & Podvoyskyi, 2018). As Ruscio (2016) wrote, using citation-based measures of merit is quick: With the calculation of one number, the evaluation is done. The LDI—the scientometric measure proposed in the following paragraphs—is similarly an easy-to-compute number. It is based on two premises. First, the more culturally diverse the experience contained in the scientific text, the higher the quality of the text. Second, as Lewin (1930) argued, to construct scientific knowledge, it is equally important to study phenomena that are common and phenomena that are rare, even if they happen only once. When applied to culture, this means that studying cultures with a 100 million members—such as English- or Chinese-speaking cultures—is equally as important as studying cultures with just one member such as the last surviving member of an Indigenous nation. Therefore, the more a culture is omitted from scientific discourse, the better the text is because it provides the experience coming from that culture—even if that culture were to consist of one last, single person. LDI should, therefore, give value to languages of small local cultures.

Current scientific evaluation practices are based on how often a work is cited, so there are several well-developed databases that track the citations of scientific works. These databases might be used for assessing the cultural diversity of a scientific text by assigning the linguistic diversity of the cited sources. To reach the premises stated above, the LDI score is computed from the sources cited in the references of the evaluated text: (a) the more linguistically diverse the sources cited in the references, the higher the LDI score, and (b) the rarer the languages cited in the references, the higher the score. The LDI score would be computed this way:

1. Once a year will be computed how many citations of publications written in various languages were used in all publications written in language A for the preceding 5 years. All cited languages in publications
written in A will be ordered from the most cited language \( \alpha_1 \) to the least cited language \( \alpha_n \). The order of languages shows how much a specific language is included in the publications in language A. Function \( \phi_i \), from the set of all languages to the set of natural numbers, is defined as giving a logarithm of order of language \( \alpha_i \) as follows: \( \phi_i(\alpha_i) = \log_{10} i \). If the first most cited language in A-written texts is, for example, English, then \( \phi_1(\text{English}) = 0 \), and the \( \phi_i \) of other languages will be nonzero.

2. Let’s have language \( \beta \) and article/publication \( p \) and let’s define \( \lambda(p, \beta) \) as a ratio:

\[
\lambda(p, \beta) = \frac{\text{Number of } p \text{'s references to sources written in the language } \beta}{\text{Number of all references in } p}.
\]

3. Let’s define \( \Psi(p) \) as follows:

\[
\Psi(p) = 1 + \log_{10} \left( \frac{\text{number of languages in which are written texts cited in } p}{\text{number of all references in } p} \right).
\]

4. The LDI(\( p \)) of article \( p \) written in language A, which cites references from languages \( \beta_1, \beta_2, \ldots, \beta_m \), is defined as follows:

\[
\text{LDI}(p) = \Psi(p) \times \left( \frac{\lambda(p, \beta_1) \times \phi_A(\beta_1) + \cdots + \lambda(p, \beta_m) \times \phi_A(\beta_m)}{\text{Number of all references in } p} \right) \tag{1}
\]

where \( \Psi(p) \) will always be lower than 3 (there would never be more than 100 languages cited and \( \log_{10} 100 = 2 \)). There would be less than 2,048 languages with written sources to be citable, so \( \phi_i(\alpha_i) \) will always be lower than \( \log_{10} 2048 = 11 \) for any pair of languages \( A, \alpha_i \). The number in the brackets on the right in Equation 1 will be always lower than 11. The LDI will therefore be a number between \( 0 \) and 33 (\( = 3 \times 11 \)).

If an article cites only publications from the most cited language in the publications written in the language of that article, the LDI for the article will be 0. The LDI of an article will be higher for publications written in more uncommon languages cited in this article (the brackets on the right) and also for publications with cited sources written in more languages, \( \Psi(p) \).

The LDI for a journal could be defined as the average of the LDIs of the articles it published in the previous year. The journal’s LDI would therefore be a number between 0 and 33. Computing such an index will allow for the ranking of journals according to their inclusion of texts that cite sources from rare and/or multiple languages.

Google Scholar might be a good citation database to calculate LDI because it is less biased toward languages other than English compared with other databases (Dahler-Larsen, 2018). Web of Science and Scopus, in 2018, contained documents written in only 51 languages (Vera-Badceta et al., 2019). The reason is that Web of Science and Scopus consist mostly of academic publications (Aksnes et al., 2019), whereas Google Scholar has a wider range of texts (Kousha & Thelwall, 2008). For a meaningful evaluation of published research according to cited languages, it might also be suitable to use a larger set of resources such as texts published in the media or on the internet (Ravenscroft et al., 2017).

Both \( \Psi \) and \( \phi \) in Equation 1 could be computed differently than proposed above. Any function that computes a value based on the number of cited languages could be used as \( \Psi \) and any function that computes a value based on the rarity of languages could be used as \( \phi \).

**Example**

Let us present an example of the computation of the LDI for the current text. First, we need an order of languages. For this purpose, we use an order of languages according to how often various languages were cited in documents written in English, indexed by Microsoft Academic, and published in the years 2015–2019. The order is therefore the same, as would be valid if Microsoft Academic were used as the database for computing LDI using the Algorithm 1 to 4 presented above. Microsoft Academic contains 11,886,288 documents published in English with publication years from 2015 to 2019. The order of languages according to how often sources written in these languages were cited in these 11,886,288 documents is presented in Table 1. There are five languages cited in the references in this text: English (46 times), Russian (3), Czech (3), Korean (1), and Polish (1). Altogether, there are 54 references. As we can see in Table 1, the rankings of these five languages are as follows: The most common language cited in English-written documents is English. Korean is the eighth most cited language in English-written documents according to Microsoft Academic, Polish is 10th, Russian is 12th, and Czech is 16th. If we use this as the order for the computation of \( \phi \) (i.e., the order or other function that is computed based on rarity), the LDI of this text would be as follows:

\[
\begin{align*}
(1 + \log_{10} 5) \times & \left( \frac{46}{54} \times \log_{10} 1 + \frac{3}{54} \times \log_{10} 12 + \frac{3}{54} \times \log_{10} 16 + \right) \\
& \left( \frac{1}{54} \times \log_{10} 8 + \frac{1}{54} \times \log_{10} 10 \right) \\
= (1 + 0.699) \times & \left( \frac{46}{54} \times 0 + \frac{3}{54} \times 3.585 + \frac{3}{54} \times 4 + \right) \\
& \left( \frac{1}{54} \times 3 + \frac{1}{54} \times 3.322 \right) \\
= 1.699 \times (0 + 0.199 + 0.222 + 0.056 + 0.062) \\
= 1.699 \times 0.539 = 0.916.
\end{align*}
\]

**Discussion**

*Explanation of LDI*

The purpose of this article is to introduce a new scientometric index—the Language Diversity Index. This index is not
based on previously published indices and should solve a problem largely omitted in scientometric scientific discourse—the language diversity of sources, the lack of which results in the researchers’ view of their research question. This is different than the issue of disciplinary diversity more often discussed by scientometric community (e.g., Zitt & Bassecoulard, 2008). The proposed index should contribute to the multidimensional evaluation of research as demanded by social science and humanities studies (Toledo, 2018). In the following paragraphs, we therefore focus on the explanation of this index and the role it should serve in scientific discourse.

Computation of the LDI of a document written in language A varies according to this language. First, how often were the publications written in various languages cited in documents written in A in the previous 5 years should be computed (this computation could be done only yearly). All cited languages are ordered according to how often they appear in the references of A-language documents. Function $\Phi_A$ then transforms each language’s order to its logarithm, which is used as a value of this language in the computation formula. This formula consists of the rarity part, which computes the rarity of languages cited in the references as a product of these languages’ value and their frequency in references, and the diversity part, which is the logarithm of the number of various languages of published documents in the references. Logarithms are used to make the final LDI value smaller. The logarithm in the diversity part has a larger base than the logarithm in the rarity part to give greater stress on the rarity of languages than on the number of languages. This should give higher value to languages of small cultural groups.

Language A could be any language in which a document is published (Korean, Czech, Swahili, English, etc.). Because the order of cited languages depends on the language A, documents with the same reference list might have different LDIs, depending on the language in which the document is published. The order of languages used in the LDI computation depends on the database from which it is computed. Different databases have different coverage of various languages documents; the order of cited languages will therefore vary according to this coverage.

Scientometric indices should lead to desirable changes in scientific discourse (Hammersley, 2014). LDI has been created to support such desirable change—to encourage researchers to get closer to the communities they research, that members of these communities might have a greater chance of being included in scientific discourse, and that information between cultures can circulate more often without the intermediate role of English scientific discourse.

The suggested index might be helpful if the following propositions are valid:

**Proposition 1:** There is at least one language for which documents published in this language do not cite sources from all languages evenly.

**Proposition 2:** Lewin’s (1930) opinion that sciences should study common phenomena as thoroughly as phenomena

| Languages ordered by number of citations in A | Number of times documents published in this language were cited in documents published in 2015–2019 in A = English |
|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------|
| $\alpha_1$ = English                        | 326,005,701                                                                                                                     |
| $\alpha_2$ = Japanese                       | 4,222,865                                                                                                                       |
| $\alpha_3$ = Spanish                        | 487,191                                                                                                                         |
| $\alpha_4$ = French                         | 407,340                                                                                                                         |
| $\alpha_5$ = German                         | 372,443                                                                                                                         |
| $\alpha_6$ = Portuguese                     | 254,774                                                                                                                         |
| $\alpha_7$ = Chinese (both simplified and traditional) | 196,326                                                                 |
| $\alpha_8$ = Korean                         | 161,571                                                                                                                         |
| $\alpha_9$ = Italian                        | 102,735                                                                                                                         |
| $\alpha_{10}$ = Polish                       | 92,331                                                                                                                          |
| $\alpha_{11}$ = Persian                     | 52,036                                                                                                                          |
| $\alpha_{12}$ = Russian                      | 46,739                                                                                                                          |
| $\alpha_{13}$ = Indonesian                  | 43,919                                                                                                                          |
| $\alpha_{14}$ = Dutch                        | 28,283                                                                                                                          |
| $\alpha_{15}$ = Turkish                      | 24,401                                                                                                                          |
| $\alpha_{16}$ = Czech                        | 14,941                                                                                                                          |
| $\alpha_{17}$ = Finnish                      | 14,423                                                                                                                          |
| $\alpha_{18}$ = Croatian                     | 12,959                                                                                                                          |
| $\alpha_{19}$ = Catalan                      | 7,717                                                                                                                            |
| $\alpha_{20}$ = Swedish                      | 7,361                                                                                                                            |
happening only once, because it cannot be known in advance study which phenomenon might bring greater increase in knowledge, is valid. Therefore, a small language environment might bring breakthrough with the same probability as a large language environment.

**Proposition 3:** Studying a phenomenon happening in the concrete cultural environment might attain a better quality when the researcher knows the language spoken in that cultural community.

**Proposition 4:** Researchers change their behavior according to the currently used scientometric indices.

**Proposition 5:** LDI benefits those researchers citing rarer languages.

Proposition 1 is valid, because languages will be never known and knowledge distributed evenly. Propositions 2 and 3 depend on the values of the researcher. Proposition 4 is supported by the fact that researchers change their behavior to increase their citation-rate-based scientometric indices (López-Navarro et al., 2015). For example, in the humanities, more publications are in English and in peer-reviewed publications, to increase their scientometric ratings (Hammarfelt, 2017). Proposition 5 is a consequence of the LDI computation algorithm. If these propositions are valid, then LDI should change scientist’s behavior to search language sources more equally across all languages, giving a more equal chance to discover scientific findings across all language environments, and this way, improve the quality of science.

**Impact of Language Diversity on Scientific Discourse**

In a monolingual, globalized scientific discourse, authors expect readers and reviewers to favor this discourse. They prefer to include references to the most internationally cited authors and remove references to the cultural heritage not shared by a globalized reader: “Authors prefer citations to recent thinkers who write in English: Austin, Grice, Searle, Wittgenstein . . . rather than Hegel, Husserl or Bergson” (Maingueneau, 2016, p. 116). This approach leads to “the impoverishment of scientific creation” (p. 116). Sharing the same global, English, scientific discourse decreases theoretical conflicts in the research community. This increases the danger of false consensus, when “researchers are (more and more as time passes) professionals who do their job and avoid challenging dominant assumptions” (Maingueneau, 2016, p. 118).

When scientific discourse is divided into more linguistically diverse communities, these multiple scientific discourses have different epistemological assumptions, allowing the creation of better science than when these discourses are merged into one in the global English discourse. The quality of science might be therefore improved when English serves not as a language of the production of knowledge but only as a language of transmission of knowledge among cultures (Maingueneau, 2016). English is the dominant language in global, scientific discourse (Stockemer & Wigginton, 2019), which negatively affects social science and humanities studies. Nevertheless, it is supported by the current scientometric practices, which face numerous critiques. Researchers in the social sciences and humanities should not rely on “laissez faire attitude and wait for these criticisms to change the new reality of science accountability as this is very unlikely to happen” (Pajic et al., 2019, p. 89). Adopting the LDI proposed in this article might be one of the active strategies that researchers might employ.

### $A = \text{English}$

Researchers in Western countries are currently not motivated to study non-Western populations because research evaluation practices favor research of Western cultures (Brady et al., 2018). This leads to “anglophone scientists neglecting foreign language publications” (Ammon, 2012, p. 338). Ammon (2012) thinks that the “major reason why publications in languages other than English do not often reach the global level is the Anglophones’ disinclination to learn foreign languages” (p. 349). Computing LDI for language A being English might therefore create an incentive to learn foreign languages and study other cultures. It might help to create an incentive for the inclusion of more diverse group of scientists who are able to ask more variable questions as required by Rad et al. (2018). The resulting research might increase its interpretive power—the ability to understand the cultural contexts of individuals’ behavior (Brady et al., 2018).

### $A = \text{Some Language Other Than English}$

National research evaluation policies in smaller countries (e.g., in Eastern Europe) force researchers to publish in journals indexed in Web of Science and abandon publishing in their own language (Pajic, 2015). If journals are evaluated only by their impact factors, it motivates scholars to publish only in English because articles written in English get a higher number of citations than those written in other languages (Di Bittetti & Ferreras, 2017). As a result, journals stop publishing in local languages and switch their language to English, which “might hamper the intermediary role of science in the society at large” (Schuermans et al., 2010, p. 422). LDI could reach higher values both for local languages and languages of other countries, so if it is used in research evaluation, it will increase researchers’ and journals’ motivation to publish in the local language. Currently, authors lose the ability to write well in their native language because of writing only in English (Nyggaard, 2019). If writing in their native language were more valued, they might improve this ability. This might help researchers to preserve the identity of the local language as a scientific language (Li, 2019).

When there is no incentive to learn and to read documents written in other foreign languages than English, researchers
might be motivated to ignore other foreign languages and cite only sources in their native language and in English because it increases their citations. As discussed by Gong et al. (2019), Chinese language texts receive more citations when they cite foreign language documents and this foreign language is nearly always English.

Several social science subdisciplines are centered on a particular cultural community and they use this community language as the communication language of their subdiscipline (Stockemer & Wigginton, 2019). LDI will motivate researchers to cite sources from smaller languages, which might increase the reputation of texts published in these languages, given that “the primary motivator of choosing English as a publication language is the belief that publishing in English will increase the reputation of one’s work” (Stockemer & Wigginton, 2019, p. 645), LDI might lead to increased publishing in smaller languages and enhancing the quality of their scientific discourse.

Research production in non-Western countries is mostly read by local scholars and does not have the opportunity to influence global science (Tijssen et al., 2006) because inclusion of local and minority cultures in global science is not encouraged by the current system of scientific evaluation, and cultural diversity is not included in its notion of the quality of science (Pontille & Torny, 2010). Publishing in English in Western journals and publishing houses leads to situations in which local communities and languages are connected to the West, but they are not connected to each other (Neylon, 2020). This diminishes international cultural exchange. LDI is computed for each language separately, so it benefits local researchers who focus on connecting their language community to communities in small countries. Because LDI values the citing of various languages, researchers from other cultures are motivated to learn this culture’s language, so texts written in this language might spread knowledge globally without the need for English to serve as mediator of knowledge. Scientific discourse written in the researcher’s language will be also enriched by the usage of LDI because researchers will be motivated to include knowledge from more diverse set of cultures into their texts. Its usage might support cultural diversity in small non-Western cultures.

**Benefits of the LDI**

Current scientific evaluation practices do not promote cultural and linguistic diversity. LDI is designed to encourage linguistic diversity and cultural diversity in published academic texts. If the linguistic diversity of sources cited in academic texts increases, it will increase the chance that opinions, ideas, and the lives of linguistic and cultural minorities (in a specific country or globally) are included in published texts.

LDI acknowledges the fact that the same knowledge does not have the same value in two different language communities. Knowledge of Burmese is rarer in the Czech Republic than in Thailand, whereas the knowledge of Polish is rarer in Thailand than in the Czech Republic. That is the reason that the order of the languages of the cited sources would be computed separately for each language in which articles are published. The usage of LDI might, therefore, encourage linguistic diversity in all academic texts in all language environments.

The humanities are heterogenous (Fanelli & Glänzel, 2013), which makes their measurement by scientometric indices more difficult. However, they are generally more dependent on the language used by the studied population than natural sciences. Computing scientometric indices from the number and rarity of languages cited in a text is more effective for judging humanities when compared with the indices that measure the number of citations at a certain time. Research in the humanities might take a long time to be cited; sometimes, it takes more than 10 years to get citations (Ardanuy et al., 2009). When research is evaluated according to the number of citations it gets in the preceding few years, texts in the humanities cannot be compared because many of them (despite being quality work) would not yet have citations. LDI overcomes this problem because all of the necessary information for its computation is already in the text.

Research evaluation practice should take into consideration differences between scientific disciplines; otherwise, it risks epistemic injustice between disciplines. Current scientometric practices are better suited for laboratory sciences than for the humanities because they serve more disciplines aimed to create facts than disciplines aimed to enhance understanding (Lohkivi et al., 2012). Evaluation practices should be improved to better conform to epistemic aims of disciplines, which creates understanding between cultures by sharing knowledge contained in sources written in those cultures’ languages. LDI might enhance the value of research, which needs sources from small and minority languages in disciplines like cultural psychology and therefore creates a more epistemic just environment for them.

LDI is designed to enhance linguistic diversity, which is not the same as cultural diversity. Nevertheless, in most of the world, separate cultural groups are usually separate linguistic groups. This is especially true for Indigenous nations in many countries. These Indigenous nations do not have their own academic literature, but if citations of nonacademic texts would be included in the computation of LDI, it would benefit the texts written in these Indigenous languages when they are being cited in academic texts. Such an inclusion might benefit those indigenous minorities.

**Limitations of the LDI**

No scientometric index is perfect. LDI has some weak points that are inevitable when it is computed from the number of languages in the references. First, the definition of a separate language can be a political, rather than an academic, issue. Serbian and Croatian are nearly the same language, but they...
are considered distinct because the countries where they are spoken are separate. On the contrary, different Chinese dialects are considered to belong to one language despite being mutually incomprehensible. Researchers who read in a language with an existing similar language that is considered separate because of political issues will have an advantage when computing LDI. Second, it would be easy to increase the LDI by citing papers that the researcher never read because they do not understand the language of these papers. However, this problem could be resolved by employing field/site visits, a research evaluation method where the evaluator speaks with the researchers about their experience and research strategies (Pedersen et al., 2020). LDI should be used together with other methods of measuring scientific quality. If not used together with other scientometric indices, LDI might lead researchers to pursue linguistic diversity at the expense of other aspects of scientific quality.

Another one of LDI’s limitations is that bibliometric databases do not cover enough documents in smaller languages (Ochsner et al., 2017). Given the current databases’ coverage, the usage of LDI might be limited by this fact. If LDI is used for the evaluation of scientific research, companies producing these databases should work on improving the coverage of smaller languages. Another problem is that some databases might not identify the document’s language with good precision. Correct identification of the document language is necessary for LDI to be effective.

**Conclusion**

Research quality is not equivalent to being cited (Feist, 2016). It consists of several dimensions such as scientific quality, plausibility, originality, and societal value. Citation indicators (i.e., indicators for how often a text is cited) are “of little help in the evaluation of the . . . societal value of research” (Aksnes et al., 2019, p. 12). One aspect of the societal value of research is the representation of ideas, values, and opinions from minority and remote cultures such as language communities whose languages are rarely mentioned in the society. There is a lack of scientometric indicators to evaluate this representation of the minority and remote language communities in scientific discourse. This article offers such an indicator to measure the rarity of cultural influence in a scientific discourse based upon the citations of the sources published in various languages.

Even if many scientists consider the quantitative measurements of scientific output impossible (Funk, 2016), some sort of measurement is inevitable because of its usability for governmental authorities. It is, therefore, worth using such measures, which support positive changes in society. Focusing on the improvement of bibliometric measures changes the behavior of scientists. Focus on being cited means that more citations are used when writing articles, and journals might require more citations (Weingart, 2005). Focusing on citing linguistically variable sources might lead to more such sources being cited by scientists and required by journals. If some journals adopt the increase of LDI as a goal, it might lead to the increase of linguistic diversity of opinions, views, and ideas cited in the published articles. This might lead to the increased presence of values and the views of indigenous, minority, and small cultural communities. Such a development might help in the production of scientific texts that help to develop a more just and better world.

**Acknowledgment**

We thank to Pavel Šmerk for giving us information from Microsoft Academic presented in Table 1.

**Author Contributions**

V.L. presented the idea, developed the algorithm for computation of LDI and the example in the article, created conceptualization of the text, wrote the first version of the manuscript, and participated in the manuscript rewriting and revisions. K.O., E.C., and G.H. participated in the manuscript rewriting and revisions. All authors have approved the final version of the article.

**Declaration of Conflicting Interests**

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

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: The work of Václav Linkov was supported by the Ministry of Education, Youth and Sports within National Sustainability Programme I, a project of Transport R&D Centre (LO1610), on a research infrastructure acquired from the Operational Programme Research and Development for Innovations (CZ.1.05/2.1.00/03.0064).

**Ethical Statement**

This article does not contain any research on human or animal subject.

**ORCID iD**

Václav Linkov https://orcid.org/0000-0001-8473-1465

**References**

Aksnes, D. W., Langfeldt, L., & Wouters, P. (2019). Citations, citation indicators, and research quality: An overview of basic concepts and theories. *SAGE Open, 9*(1). https://doi.org/10.1177/2158244019829575

Ammon, U. (2012). Linguistic inequality and its effects on participation in scientific discourse and on global knowledge accumulation—With a closer look at the problems of the second-rank language communities. *Applied Linguistics Review, 3*, 333–355. https://doi.org/10.1515/applrev-2012-0016

Ardanuy, J., Urbano, C., & Quintana, L. (2009). A citation analysis of Catalan literary studies (1974–2003): Towards a bibliometrics of humanities studies in minority languages. *Scientometrics, 81*, 347–366.
Arnett, J. J. (2008). The neglected 95%: Why American psychology needs to become less American. American Psychologist, 63, 602–614. https://doi.org/10.1037/0003-066X

Barr, R. (2019). Les indicateurs sont morts, vive les indicateurs! Towards a political economy of SandT indicators: A critical overview of the past 35 years. Research Evaluation, 28, 2–6. https://doi.org/10.1093/reseval/rvy029

Brady, L. M., Fryberg, S. A., & Shoda, Y. (2018). Expanding the interpretive power of psychological science by attending to culture. Proceedings of the National Academy of Sciences of the United States of America, 115(45), 11406–11413. https://doi.org/10.1073/pnas.1803526115

Chavarro, D., Rafols, I., & Tang, P. (2018). To what extent is inclusion in the Web of Science an indicator of journal “quality”? Research Evaluation, 27, 106–118. https://doi.org/10.1093/reseval/rvy001

Choi, S. H., & Han, G. (2000). Munhwashimirihaekjok yokunbang-poplon [Methodology of cultural psychology]. Korean Journal of Social and Personality Psychology, 14, 123–144.

Dahler-Larsen, P. (2018). Making citations of publications in languages other than English visible: On the feasibility of a PLOTEx-index. Research Evaluation, 27, 212–221.

Di Bittetti, M. S., & Ferreras, J. A. (2017). Publish (in English) or perish: The effect on citation rate of using languages other than English in scientific publications. Ambio, 46, 121–127. https://doi.org/10.1007/s13280-016-0820-7

Diekhoff, T., Schlattmann, P., & Dewey, M. (2013). Impact of Article language in multi-language medical journals—A bibliometric analysis of self-citations and impact factor. PLOS ONE, 8(10). Article e76816. https://doi.org/10.1371/journal.pone.0076816

Dinkel, A., Berth, H., Borkenhagen, A., & Brähler, E. (2004). On raising the international dissemination of German research: Does changing publication language to English attract foreign authors to publish in a German basic psychology research journal? Experimental Psychology, 51, 319–328. https://doi.org/10.1027/1618-3169.51.4.319

Donovan, C. (2008). The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research. New Directions for Evaluation, 118, 47–60.

Fanelli, D., & Glänzel, W. (2013). Bibliometric Evidence for a Hierarchy of the Sciences. PLOS ONE, 8(6), Article e66938. https://doi.org/10.1371/journal.pone.0066938

Feist, G. J. (2016). Intrinsic and extrinsic science: A dialectic of scientific fame. Perspectives on Psychological Science, 11, 893–898.

Filippov, I. S. (2016). Kak ucheniye gumanitarnogo profilya otsenivayut naukometriyu [How scientists in the humanities assess scientometrics]. Siberian Historical Research, 2016/3, 6–27. https://doi.org/10.17223/2312461X/13/1

Funk, D. A. (2016). Naukometriya v otsenke kachestva publikatsiy v sotsialnykh i gumanitarnikh naukakh [Scientometrics and evaluation of publications in social sciences and humanities]. Siberian Historical Research, 2016/1, 8–26. https://doi.org/10.17223/2312461X/11/2

Furnham, A. (2020). What I have learned from my Google Scholar and H Index. Scientometrics, 122, 1249–1254.

Gong, K., Xie, J., Cheng, Y., Lariviere, V., & Sugimoto, C. R. (2019). The citation advantage of foreign language references for Chinese social science papers. Scientometrics, 120, 1439–1460.

Hammarfelt, B. (2017). Four claims on research assessment and metric use in the humanities. Bulletin of the Association for Information Science and Technology, 43(6), 33–38.

Hammersley, M. (2014). The perils of “impact” for academic social science. Contemporary Social Science: Journal of the Academy of Social Sciences, 9, 345–355.

Hoblik, J. (2020). Hodnota hodnocení: K mechanismům akademické hodnocení [Value of evaluation: On academic evaluating mechanisms]. ERGOT. Revue Pro Filosofii a Společenské Vědy, 2020(1). https://ergotisite.wordpress.com/2020/11/26/hodnota-hodnoceni-k-mechanismum-akademickeho-hodnoceni/

Kousha, K., & Thelwall, M. (2008). Sources of Google Scholar citations outside the Science Citation Index: A comparison between four science disciplines. Scientometrics, 74, 273–294. https://doi.org/10.1007/s11192-008-0217-x

Kuleshova, A. V., & Podvoyskyi, D. G. (2018). Paradoksy publikationoi aktivnosti v pole sovremennoi rossiiskoi nauki: genesi, diagnoz, trendy [Paradoxes of Publication Activity in the Field of Contemporary Russian Science: Genesis, Diagnosis, Trends]. Monitoring of Public Opinion: Economic and Social Changes, 2018/4, 169–210.

Lewin, K. (1930). The conflict between Aristotelian and Galileian modes of thought in contemporary psychology. Journal of General Psychology, 5, 141–177.

Li, X. (2019). What is the publication language in humanities? The case of Translation Studies scholars. English Today, 35(2), 42–49. https://doi.org/10.1017/S0266078418000202

Linkov, V. (2014). Několik poznámek k vytváření interkulturních psychologických výzkumů [Few notes on conducting cultural comparisons in psychological research]. Psychologie a Jeji Kontexty, 5/2, 101–108.

Lokkivi, E., Velbaum, K., & Eigi, J. (2012). Epistemic injustice in research evaluation: A cultural analysis of the humanities and physics in Estonia. Studia Philosophica Estonica, 52, 108–132. https://doi.org/10.12697/spe.2013.6.2.05

López-Navarro, I., Moreno, A. I., Quintanilla, M. Á., & Rey-Rocha, J. (2015). Why do I publish research articles in English instead of my own language? Differences in Spanish researchers’ motivations across scientific domains. Scientometrics, 103, 939–976. https://doi.org/10.1007/s11192-015-1570-1

Mainguenau, D. (2016). Monolingualism and creativity. Scientific discourse and linguistic diversity in human and social sciences. MÉTODE Science Studies Journal, 6, 115–119. https://doi.org/10.7203/metode.6.4552

Meneghini, R., Packer, A. L., & Nassi-Caló, L. (2008). Articles by Latin American Authors in prestigious journals have fewer citations. PLOS ONE, 3/11, Article e3804. https://doi.org/10.1371/journal.pone.0003804

Neylon, C. (2020). Research excellence is a neo-colonial agenda (and what might be done about it). In E. Kraemer-Mbula, R. Tijssen, M. L. Wallace, & R. McLean (Eds.), Transforming research excellence (pp. 92–115). Cape Town: African Minds, New Ideas from the Global South.

Nygaard, L. P. (2019). The institutional context of “linguistic injustice”: Norwegian social scientists and situated multilingualism. Publications, 7, Article 10. https://doi.org/10.3390/publications7010010

Ochsner, M., Hug, S., & Galleron, I. (2017). The future of research assessment in the humanities: Bottom-up assessment procedures. Palgrave Communications, 3, Article 1720. https://doi.org/10.1057/palcomms.2017.20
Ochsner, M., Hug, S. E., & Daniel, H.-D. (2013). Four types of research in the humanities: Setting the stage for research quality criteria in the humanities. *Research Evaluation, 22*, 79–92. https://doi.org/10.1093/reseval/rvs039

Pajic, D. (2015). Globalization of the social sciences in Eastern Europe: Genuine breakthrough or a slippery slope of the research evaluation practice? *Scientometrics, 102*, 2131–2150. https://doi.org/10.1007/s11192-014-1510-5

Pajic, D., Jevremov, T., & Škoric, M. (2019). Publication and citation patterns in the social sciences and humanities: A national perspective. *Canadian Journal of Sociology, 44*, 67–94. https://doi.org/10.29173/cjs29214

Pedersen, D. B., Grønvad, J. F., & Hvidtfeldt, R. (2020). Methods for mapping the impact of social sciences and humanities—A literature review. *Research Evaluation, 29*, 4–21.

Peterson, D. (2017). The depth of fields: Managing focus in the epistemic subcultures of mind and brain science. *Social Studies of Science, 47/1*, 53–74.

Pontille, D., & Torny, D. (2010). The controversial policies of journal ratings: Evaluating social sciences and humanities. *Research Evaluation, 19/5*, 347–360.

Rad, M. S., Martingano, A. J., & Ginges, J. (2018). Toward a psychology of Homo sapiens: Making psychological science more representative of the human population. *Proceedings of the National Academy of Sciences of the United States of America, 115*, 11401–11405. https://doi.org/10.1073/pnas.1721165115

Ravenscroft, J., Liakata, M., Clare, A., & Duma, D. (2017). Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements. *PLOS ONE, 12/3*, Article e0173152. https://doi.org/10.1371/journal.pone.0173152

Reale, E., Avramov, D., Canhial, K., Donovan, C., Flecha, R., Holm, P., . . . Van Horik, R. (2018). A review of literature on evaluating the scientific, social and political impact of social sciences and humanities research. *Research Evaluation, 27*, 298–308.

Ruscio, J. (2016). Taking advantage of citation measures of scholarly impact: Hip hip h index! *Perspectives on Psychological Science, 11*, 905–908.

Schuermans, N., Meeus, B., & De Maesschalck, F. (2010). Is there a world beyond the Web of Science? Publication practices outside the heartland of academic geography. *Area, 42/4*, 417–424. https://doi.org/10.1111/j.1475-4762.2010.00938.x

Stockemer, D., & Wigginton, M. J. (2019). Publishing in English or another language: An inclusive study of scholar’s language publication preferences in the natural, social and interdisciplinary sciences. *Scientometrics, 118*, 645–652. https://doi.org/10.1007/s11192-018-2987-0

Strehl, L., Calabró, L., Souza, D. O., & Amaral, L. (2016). Brazilian science between national and foreign journals: Methodology for analyzing the production and impact in emerging scientific communities. *PLOS ONE, 11/5*, Article e0155148. https://doi.org/10.1371/journal.pone.0155148

Tijssen, R. J. W., Mouton, J., van Leeuwen, T. N., & Boshoff, N. (2006). How relevant are local scholarly journals in global science? A case study of South Africa. *Research Evaluation, 15/3*, 163–174.

Toledo, E. G. (2018). Research assessment in humanities and social sciences in review. *Revista Española de Documentación Científica, 41*(3), Article e208. https://doi.org/10.3989/redc.2018.3.1552

Towpik, E. (2015). IF-mania: *Journal Impact Factor* nie jest właściwym wskaźnikiem oceniania wyników badań naukowych, indywidualnych uczonych ani ośrodków badawczych [IF-mania: *Journal Impact Factor* is not a proper mean to assess the quality of research, individual researchers, nor scientific institutions]. *NOWOTWORY Journal of Oncology, 65*, 465–475.

Urbánek, T., & Čeněk, J. (2019). Adaptace a ekvivalence testových metod: inspirace pro psychologické testování minorit v ČR [The adaptation and equivalence of test methods: An inspiration for psychological assessment of minorities in the Czech Republic]. *Československá Psychologie, 63/1*, 142–154.

Vera-Bacdeta, M.-A., Thellwall, M., & Kousha, K. (2019). Web of Science and Scopus language coverage. *Scientometrics, 121*, 1803–1813. https://doi.org/10.1007/s11192-019-03264-z

Weingart, P. (2005). Impact of bibliometrics upon the science system: Inadvertent consequences. *Scientometrics, 62*, 117–131.

Zitt, M., & Bassecquard, E. (2008). Challenges for scientometric indicators: Data demining, knowledge-flow measurements and diversity issues. *Ethics in Science and Environmental Politics, 8*, 49–60. https://doi.org/10.3354/esep00092