Science, Maddá and 'Ilm: Differences in the Quality of Scientific Information Available to Internet Users in Three Languages

Kawther Zoubi
Technion –
Israel Institute of Technology
skawtherz@gmail.com

Aviv J. Sharon
Technion –
Israel Institute of Technology
aviv.sharon@gmail.com

Eyal Nitzany
Independent Scholar
eyalni@gmail.com

Ayelet Baram Tsabari
Technion – Israel Institute of Technology
ayelet@technion.ac.il

סַיְיֶנְס, מַדָּע וְעִלְם: הבדלים באיכות המידע המדעי הזמין למשתמשי האינטרנט בשלוש שפות

כאותר זועבי
הטכניון – מכון טכנולוגי לישראל
skawtherz@gmail.com

אביב שרון
הטכניון – מכון טכנולוגי לישראל
aviv.sharon@gmail.com

אייל ניצני
חוקר עצמאי
eyalni@gmail.com

אילת ברעם-צברי
הטכניון – מכון טכנולוגי לישראל
ayelet@technion.ac.il

Abstract

The internet is a source of information with potential to alleviate inequality in general and specifically with respect to science literacy. Nevertheless, digital divides persist in online access and use and in subsequent social outcomes. Among these, the "language divide" partly determines how successful users are in their internet use depending on their proficiency in languages, and especially in English. To examine whether the quality of online scientific information varies between languages, we compared online search results regarding school science in English, Hebrew and Arabic. We compiled a list of thirty school science terms in each language, spanning three fields: physics, chemistry and biology. Search results were collected from *Google Search* and the quality of the first seven relevant results was evaluated (n_{total} = 630). Findings indicate that searches in English yielded overall higher quality results, compared with Hebrew and Arabic, but mostly in pedagogical aspects, rather than scientific ones. Clustering the results by language yielded better separation than clustering by scientific field, pointing to a "language divide" in access to online school science content among students within the same country. We argue that the scientific and educational communities should act to mitigate this language divide.

Keywords: Science literacy, science communication, digital divide, digital inequality, language divide.

Nur and Talia are fifth-grade students who attend an after-school science class together in Haifa, Israel. Nur is a native Arabic speaker who is proficient in Hebrew and English, and Talia is bilingual in Hebrew and English. One day, they are asked by their instructor to search for information about the processed foods they eat every day. A quick search reveals that in-depth educational videos about baking are only available in English, whereas the top search results in Hebrew and Arabic typically offer technical information for professionals. As they run more searches, they get a stronger impression that the usefulness of results differs depending on the language they search in.

This fictional vignette reflects a broader issue: Science literacy is considered to benefit the health and well-being of individuals, communities and society (National Academies of Sciences Engineering and Medicine, 2016). As access to the internet increases globally, it has the potential to alleviate social inequalities, for example, by increasing access to useful scientific information. However, legacy inequalities remain with us within and between nations of the world, in part due to differences in language proficiency. In the following section, we situate this issue within existing theories and findings.

Literature Review

The Internet as a Source of Scientific Information

The internet is a major source of information about science and technology (S&T) in developed countries. As of 2018, 57% of US adults cite the internet as their primary source of S&T information, and 70% say they would go online to find information about a specific S&T issue. (National Science Board, 2020). Similarly, in Israel, 77% of adults who mentioned that they were interested in least one field of S&T cited search engines as a primary source of S&T information (Israel Ministry of Science Technology and Space, 2017). Interest in health-related issues in both countries is usually higher than science-related issues, e.g., space exploration (US: 56 vs. 25%; Israel: 61 vs. 30%).

The Digital Divide and the Language Divide

Unfortunately, not everyone equally benefits from access to information on the internet, S&T-related or otherwise, due to disparities collectively named "the digital divide." This term refers to "any divide or gap between people […] in their communication technology awareness, adoption or ownership, use, and skill" (Pearce & Rice, 2014). Research on the digital divide has focused on three topics: physical access (e.g., in terms of hardware and connectivity; the "first-level" digital divide), use (the "second-level" digital divide), and outcomes (the "third-level"), such as health and educational outcomes (Hargittai & Hsieh, 2013; Robinson et al., 2020).

The literature shows that as internet access increases worldwide, individuals from higher socio-economic strata tend to benefit from it more than others, since they tend to possess higher levels of skill and social support (OECD, 2015). Similarly, the digital divide is also associated with other social inequalities along lines of "gender, sexuality, race and ethnicity, aging, disability, healthcare, education, rural residency" and more (Robinson et al., 2020, p. 1). Thus, in many countries there are large disparities in early exposure to computers depending on socio-economic status and gender (OECD, 2015). Moreover, a global digital divide is associated with disparities in countries' wealth, political systems, telecommunication policies and more (Hargittai & Hsieh, 2013). These digital inequalities have been exacerbated by the COVID-19 pandemic, as the "digitally disadvantaged" are less able to take advantage of eHealth services and remote learning (Robinson et al., 2020).

Here, we focus on a relatively understudied aspect of the digital divide: the "language divide" in internet adoption and use (De Jesus & Xiao, 2012), which derives from dominance of a small number of languages on the internet, mainly English. This has long been theorized as a barrier to internet adoption and use in a linguistically diverse world (Chen & Wellman, 2004; Warschauer, 2002). Relatively few studies investigated this issue, but they indicate that internet use is correlated with English proficiency in diverse contexts, including Italy and India (reviewed in Pearce & Rice, 2014) and among the Hispanic population in the US (De Jesus & Xiao, 2012).

The Role of Social Structures in Shaping Science Literacy

The sociologically-oriented conception of the "digital divide" aligns with recent conceptions of science literacy. This term has historically focused on science literacy as an individual characteristic, i.e., a person's ability to reason with and about science. However, the concept now refers to science literacy as a characteristic of communities and societies as well. A recent consensus report theorizes that structural factors can support or constrain individual's science literacy. Similarly, along the lines of the second-level digital divide, the report expresses concern that inequalities in science literacy are exacerbated due to "differences in the way that people are supported in their use of Internet technologies" (National Academies of Sciences Engineering and Medicine, 2016, p. 107).

Additionally, scholars have pointed out a language divide in science and argued that English serves as a "gatekeeper to scientific discourse" to the detriment of public communication of S&T in other languages (Márquez & Porras, 2020, p. 5). Arguably, this structural factor shapes inequality in science literacy as well.

Summary

Multiple sources suggest that the "language divide" can shape the use of online S&T information and constrain the development of science literacy, but this topic has been relatively understudied. Additionally, the public's needs and interests in scientific topics vary by topic (e.g., with respect to health vs. space exploration). Hence, there is both a theoretical and a practical motivation to understand quality of online scientific information available to users in different languages and across different fields.

Research Goal and Questions

Our goal is to examine the characteristics and quality of online scientific information and compare them across three languages: English, Hebrew, and Arabic, and across three fields (disciplines): physics, chemistry, and biology. Specifically we ask:

How does the quality of online scientific information concerning core concepts in biology, chemistry and physics differ when comparing languages and when conducting searches from the same country?

Research Context

This study focuses on scientific content in the Hebrew and Arabic languages, compared with content available in English, the dominant language of the internet (Pearce & Rice, 2014). Modern Hebrew is the official language of Israel and 49% of its population over 20 years old speaks it natively (approximately 4.5 million native speakers), with most of the rest of the population proficient in Hebrew. Most native Hebrew speakers are Jewish citizens of Israel. By contrast, Arabic has semi-official status in Israel with a large minority of native speakers (18% of the population over 20 years old, approximately 1.8 million people). Most of these are Arab citizens of Israel (Israel Central Bureau of Statistics, 2013).

While Arabic is a minority language in Israel, it is an official language in 27 other countries and is spoken by roughly 274 million people worldwide (Eberhard, Simons, & Fennig, 2020); Arab countries have been relatively late adopters of the internet (Warf & Vincent, 2007) and rates of internet usage still vary considerably between them.

Several studies point at the existence of a second-level digital divide between Jews and Arabs in Israel. The PIAAC study found that 34% of Arabs aged 16-65 have poor proficiency in accessing, analyzing and communicating information using common computer applications, compared with 9% of Jews in the same age range (Israel Central Bureau of Statistics and Israel National Authority for Measurement and Evaluation in Education, 2016). Arab internet users also report that they use the internet to search for information less often than Jewish users (46% vs. 79%, respectively; Lissitsa, 2015). Additionally, within Arab society in Israel, Hebrew and English proficiency correlates with capital-enhancing uses of the internet, such as searching for information (Lissitsa, 2015). In surveys from 2011-2014, between 61 and 68 percent of Arab surfers reported that they prefer reading Arabic-language websites, whereas 25 to 28 percent preferred Hebrew-language websites (Ganayem, 2018).

Methods

Sampling Search Terms

To measure the quality of scientific information online, a list of scientific terms in three languages was constructed in four steps: (1) Collection of core scientific terms from school science curricula and from relevant research literature; (2) Validation using a panel of secondary school science teachers; (3) Translation to English and Arabic; and (4) Refinement based on the search results.

First, we collected 365 terms in Hebrew from several sources, including secondary-school physics, chemistry and biology curricula and science content standards from the United States, Israel, and Egypt. We also included terms from scholarly articles about children's interest in science (Baram-Tsabari & Yarden, 2005) and about public engagement with science online (Segev & Sharon, 2016).

Second, for the validation step, we assembled a panel of nine secondary school science teachers, all native Arabic speakers with professional working proficiency in Hebrew and English. The panel consisted of three smaller panels of three teachers each, for physics, chemistry and biology. Each panelist held at least a bachelor's degree in a scientific discipline or in science teaching, and most (seven out of nine) held an advanced degree as well. Additionally, each panelist had at least ten years' teaching experience. The panel members were asked to select the ten most central terms to their scientific domain derived from the list generated in the previous step, with special preference to terms that they considered relevant to everyday life. The panel discussions yielded a list of 30 terms, consisting of ten terms from each scientific domain (Table 1, Hebrew column).

Third, we translated the 30 terms to English and Arabic. Since translation often yielded several possibilities, the translations were validated using the multilingual online encyclopedia, Wikipedia. The Hebrew terms were entered into the Hebrew-language edition of Wikipedia, and then equivalent terms in English and Arabic were chosen using the interlanguage links, which point from one article to its equivalent articles in other editions of the encyclopedia. Arabic translations were also validated with the teacher panels to verify alignment with common usage among Arabic speakers in Israel and within the Arabic version of the Israeli school science curriculum. Hence, for example, the term selected for "pH" was *darajat al-ḥumūḍa* (درجة الحموضة, "acidity level") rather than the term used in the Arabic Wikipedia article title, *us hīdrūjīnī* (أس هيدروجيني, "power [exponent] of hydrogen").

Fourth, a final refinement step was conducted to improve the relevance of the search results. For example, the term "volume" in English yielded results referring to both three-dimensional space and to sound pressure; the term *láẖats* in Hebrew (לחץ, "pressure") yielded results relating to psychological stress; and the term *makhlūṭ* in Arabic (مخلوط, "mixture") yielded results relating to spice mixes and certain food dishes. Thus, if at least four out of the top seven results did not relate to the scientific aspect of the term, the names of the scientific domains ("physics," "chemistry," "biology") were added to the search term in parentheses to obtain more relevant results. For example, the search term "time" was substituted with "time (physics)." This change was done for five search terms in all three languages (time, volume, mixture, pressure, and cell; Table 1, items 2, 12, 18, 19, 24).

Limitations. The reliance on Wikipedia led to some slightly different translations to English than anticipated, such as item 4 appearing in English as "electrical network" rather than the common term "electrical circuit," which refers to just one type of network. Similarly, users searching for the Hebrew article for *maẖalá torashtít* (מחלה תורשתית, "hereditary disease;" item 28) were redirected to the article titled *pgam genéti* (פגם גנטי, "genetic disorder"). Hence, the English term "genetic disorder" was included in the sample, rather than the direct translation, "hereditary disease."

Table 1. List of scientific concepts in three languages and three fields.

Field	Item No.	English	Hebrew	Arabic
Physics	1	Electrical Insulator	מבודד חשמלי	عازل كهربائي
	2	Time (Physics)*	זמן (פיזיקה)*	زمن (فيزياء)*
	3	Voltage	מתח חשמלי	جهد كهربائي
	4	Electrical network	מעגל חשמלי	دائرة كهربائية
	5	X-ray	קרינת רנטגן	أشعة سينية
	6	Light spectrum	ספקטרום האור	طيف الضوء
	7	Gravity	כבידה	جاذبية
	8	Density	צפיפות	كثافة
	9	Radiation	קרינה	اشعاع
	10	Velocity	מהירות ממוצעת	سرعة متجهة
Chemistry	11	Mass	מסה	كتلة
	12	Volume (chemistry)*	נפח (כימיה)*	حجم (كيمياء)*
	13	State of matter	מצב צבירה	حالة المادة
	14	Gas	גזים	غاز
	15	Liquid	נוזלים	سائل
	16	Chemical elements	יסודות כימיים	العناصر الكيميائية
	17	pH	רמת חומציות	درجة الحموضة
	18	Mixture (Chemistry)*	תערובת (כימיה)*	مخلوط (كيمياء)*
	19	Pressure (Chemistry)*	לחץ (כימיה)*	ضغط (كيمياء)*
	20	Ozone	אוזון	أوزون
Biology	21	Carbohydrate	פחמימה	سكريات
	22	Fat	שומן	دهن
	23	Protein	חלבון	بروتين
	24	Cell (biology)*	תא (ביולוגיה)*	خلية (علم الأحياء)*
	25	Homeostasis	הומיאוסטזיס	اتزان بدني
	26	DNA	דנ״א	د.ن.أ
	27	Metabolism	מטבוליזם	أيض
	28	Genetic disorder	מחלה תורשתית	مرض وراثي
	29	Enzyme	אנזים	انزيم
	30	Menstrual cycle	המחזור החודשי	الدورة الشهرية

Data Collection and Analysis

The search terms were entered into *Google Search* from the same computer using an Israel-based internet connection in December 2018. In total, 630 results were obtained (30 terms × 3 languages × 7 results = 630 results). We took measures to avoid surveillance that could personalize the results, including using the browser in a private browsing mode; disabling Google Search customization settings; and deleting browser history before each search.

The scientific relevance of the first seven results was determined and recorded (Table 2, row 1); if all these results pertained the scientific aspect of the term, they were included in the sample and analyzed (rows 2-12). This occurred in 66 of the 90 searches (73.3%). For the rest of the searches, any irrelevant results were disregarded. Subsequent *relevant* results were included instead, until 7 results were reached per term.

The results were coded using a codebook developed based on 40 sources on evaluating electronic information quality in general and in specific domains such as health and nutrition. Some common variables include accuracy of the content, frequency of updates and maintenance, and whether the content is freely accessible (e.g., Guardiola-Wanden-Berghe, Gil-Pérez, Sanz-Valero, & Wanden-Berghe, 2011; Savolainen, 2011; Shahbazi, Farajpahlou, Osareh, & Rahimi, 2019).

To assess inter-rater reliability, the first author and a research assistant independently coded a sub-sample of 9.5% of search results (n = 60). Cohen's Kappa values were over 0.9 for all but three variables: "Coverage" ($\kappa > 0.7$), "Everyday Life" ($\kappa > 0.8$) and "Last Updated" ($\kappa > 0.8$).

Table 2. Codebook for assessing the quality of scientific information online.

	Variable	**Description**	**Range of Possible Values**	**Range of Observed Values**
	A. Scientific Quality			
1.	Scientific Results*	Number of scientific results among the top 7 results	[4, 7]	[5, 7]
2.	Accuracy#	The extent to which the content is free of scientific errors and imprecision	[-1, 2]	[-0.57, 2]
3.	Coverage#	The comprehensiveness of the explanation	[0, 4]	[0.28, 4]
4.	No. of Sources Cited#	The number of sources cited, where all values greater than one were recoded as one	0, 1	[0, 1]
5.	Authority#	Author's education and expertise are relevant to the domain	[0, 3]	[0, 2.43]

	Variable	Description	Range of Possible Values	Range of Observed Values
	B. Pedagogical Quality			
6.	Educational Results#	Search result is an educational website designed for students	0 (No), 1 (Yes)	[0, 0.85]
7.	Everyday Life#	References to everyday life	0 (No), 1 (Yes)	[0, 1]
8.	Illustration Rating#	Use of relevant audio and/or visual materials, including animations, simulations, and video	0 (None), 1 (Illustrative or unrelated items), 2 (One relevant item), 3 (Two or more relevant items)	[0, 3]
9.	Links to New Concepts#	Search result contains links to new concepts or defines them	0 (No), 1 (Yes)	[0.14, 1]
10.	Further Reading#	Search result contains references for further reading	0 (No), 1 (Yes)	[0.14, 1]
	C. Variables Specific to Online Content			
11.	Last Updated#	The time that has passed since the content was created, in years, where items older than 3 years old were coded as 3 years old	[0, 3]	[0.14, 2.71]
12.	Interactivity#	Availability of options to contact the author(s), especially online	[0, 2]	[0, 0.57]

Note. * Measured per search term; # Measured per search result, values averaged per search term

We conducted one-way ANOVAs, comparing the means for each variable separately between languages ($n_{\text{English}} = n_{\text{Hebrew}} = n_{\text{Arabic}} = 30$) and between fields ($n_{\text{Physics}} = n_{\text{Chemistry}} = n_{\text{Biology}} = 30$). To uncover specific differences between the means, these analyses were followed up by Tukey post-hoc tests (for comparisons that met the assumption of homogeneity of variances) or by Games-Howell post-hoc tests (for the rest).

Then we conducted a Linear Discriminant Analysis (LDA) to reduce the dimensionality of input from 12 (# of variables) to 2 (x,y) by projecting it to the most discriminative directions. In our case, LDA receives 12 characteristics for each search term (e.g., accuracy and coverage ratings, etc. for "enzyme") along with a designation of its group (e.g., "English") and attempts to find a linear transformation that would plot the search terms within each given group closely together on a two-dimensional plane while maximizing the separability between the given groups. To accomplish this, the algorithm calculates two axes, or linear discriminants, LD1 and LD2, that are each correlated with sets of the input variables. Each point (search term) receives scores along these axes. The LDA was run twice: once attempting to separate search terms by languages and once by fields.

Findings and Discussion

Are there differences in quality between languages and fields?

Overall, scientific quality was similar across the three languages, except for authority ratings and source citations (Table 3A). The average search yielded between six and seven relevant results among the top seven results on average, irrespective of language. Scientific quality, accuracy and coverage were similar between languages. However, English results had the highest authority ratings ($p < .001$) and Hebrew results cited the fewest sources ($p < .001$).

By contrast, pedagogical quality and variables specific to online content differed between the languages in many ways (Table 3B). English-language results had a consistently high pedagogical quality, and they were higher than Hebrew and Arabic results with respect to links to everyday life ($p < .05$) and illustration ratings ($p < .01$); however, English did have some weaknesses compared with Hebrew and Arabic. Hebrew results were the most recent ($p < .01$) and interactive ones ($p < .001$), and Arabic the most references for further reading. Interestingly, Arabic had the fewest educational results ($p < .05$) and Hebrew results had the fewest links to new concepts ($p < .01$).

With respect to fields, a much more uniform picture emerged. The average search yielded between six and seven relevant results on average irrespective of field (Table 3A). However, overall, coverage and accuracy were found to be significantly lower for chemistry search results than those of other fields (Coverage: $p < .001$; Accuracy: $p < .05$). Authority was also found to be higher in biology results than in physics results ($p < .05$). Pedagogical quality and other variables were overall similar across fields (Table 3B & 3C) except for references to everyday life, which were less abundant – again – in chemistry results ($p < .01$).

Table 3. Information quality by language and by field

		Comparison by Language				Comparison by Field			
		Language	***M***	***SD***	**Sig.**	**Field**	***M***	***SD***	**Sig.**
	A. Scientific Quality								
1.	Scientific Results	English	6.47	0.82	-	Physics	6.60	0.77	-
		Hebrew	6.53	0.78		Chemistry	6.37	0.81	
		Arabic	6.73	0.58		Biology	6.77	0.57	
2.	Accuracy	English	1.90	0.27	-	Physics	1.90	0.19	P>C * B>C **
		Hebrew	1.79	0.38		Chemistry	1.49	0.77	
		Arabic	1.66	0.73		Biology	1.96	0.09	

		Comparison by Language				Comparison by Field			
		Language	***M***	***SD***	**Sig.**	**Field**	***M***	***SD***	**Sig.**
3.	Coverage	English	2.66	0.86	-	Physics	2.60	0.84	P>C *** B>C ***
		Hebrew	2.19	0.97		Chemistry	1.53	0.73	
		Arabic	2.11	0.95		Biology	2.82	0.73	
4.	No. of Sources Cited	English	0.97	0.13	E>H *** A>H ***	Physics	0.60	0.46	-
		Hebrew	0.17	0.32		Chemistry	0.69	0.44	
		Arabic	0.85	0.27		Biology	0.70	0.42	
5.	Authority	English	1.50	0.45	E>H *** E>A ***	Physics	0.83	0.60	B>P *
		Hebrew	0.68	0.49		Chemistry	0.92	0.49	
		Arabic	0.77	0.50		Biology	1.20	0.64	
	B. Pedagogical Quality								
6.	Educational Results	English	0.31	0.24	E>A * H>A ***	Physics	0.31	0.22	-
		Hebrew	0.38	0.19		Chemistry	0.26	0.22	
		Arabic	0.16	0.19		Biology	0.27	0.25	
7.	Everyday Life	English	0.84	0.19	E>H * E>A ***	Physics	0.79	0.19	P>C *** B>C **
		Hebrew	0.67	0.31		Chemistry	0.52	0.28	
		Arabic	0.55	0.25		Biology	0.75	0.28	
8.	Illustration Rating	English	2.12	0.32	E>H *** E>A ** A>H ***	Physics	1.86	0.50	-
		Hebrew	1.26	0.58		Chemistry	1.54	0.69	
		Arabic	1.79	0.32		Biology	1.76	0.39	
9.	Links to New Concepts	English	0.58	0.20	E>H ** A>H ***	Physics	0.57	0.24	-
		Hebrew	0.38	0.19		Chemistry	0.50	0.21	
		Arabic	0.66	0.20		Biology	0.54	0.23	
10.	Further Reading	English	0.75	0.17	E>H *** A>E ** A>H ***	Physics	0.74	0.21	-
		Hebrew	0.55	0.20		Chemistry	0.70	0.23	
		Arabic	0.87	0.14		Biology	0.72	0.22	

		Comparison by Language				Comparison by Field			
		Language	***M***	***SD***	**Sig.**	**Field**	***M***	***SD***	**Sig.**
	C. Variables Specific to Online Content								
11.	Last Updated (Years)	English	1.78	0.49	E>H *** A>H **	Physics	1.41	0.52	-
		Hebrew	1.17	0.33		Chemistry	1.55	0.54	
		Arabic	1.54	0.65		Biology	1.52	0.62	
12.	Interactivity	English	0.12	0.12	H>E *** H>A ***	Physics	0.16	0.14	-
		Hebrew	0.27	0.14		Chemistry	0.16	0.16	
		Arabic	0.08	0.11		Biology	0.14	0.16	

Note. Sig.: Significant differences. E: English; H: Hebrew; A: Arabic; P: Physics; C: Chemistry; B: Biology; * $p < .05$; ** $p < .01$; *** $p < .001$.

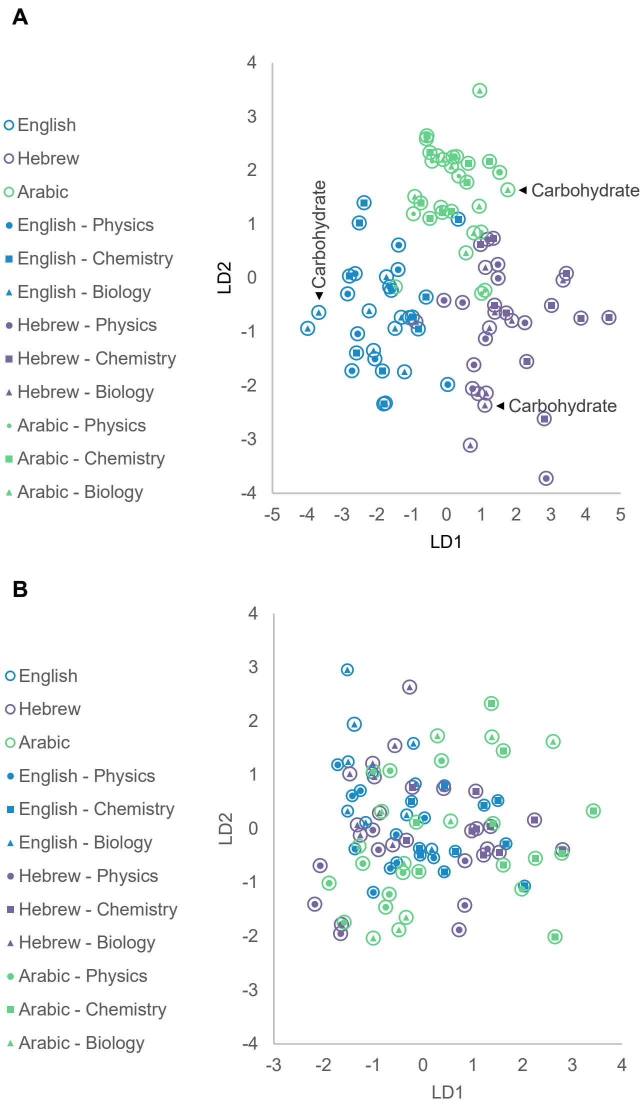
Do terms cluster together better by language or by field?

The results of the LDAs show that when applying the clustering algorithm by language, three almost separate clusters emerged, with only a few points overlapping, indicating that the search terms in each language share a set of characteristics in common that differs from the other languages (Figure 1A). By contrast, when attempting to cluster the data by field, the clusters overlapped much more, indicating that it is more difficult to find common characteristics within each of the scientific fields (Figure 1B).

Which terms have the most similar characteristics across languages and which terms differ the most?

To measure the distances between equivalent terms in different languages, we calculated the areas of the triangles determined by the three data points representing equivalent terms in Figure 1A (e.g., between the three data points representing "carbohydrate," when clustering the data by language). This analysis shows that the list of 15 most similar terms across languages is mostly composed of chemistry and physics terms, with 7 and 5 terms respectively (Figure 1C). Nevertheless, the "most similar" term overall is "menstrual cycle" (biology).

Conversely, the 15 most dissimilar terms are made up of mostly biological terms, with some of the most dissimilar terms relating to biochemistry and nutrition ("carbohydrate," "protein," "enzyme" and "metabolism"). Other highly dissimilar terms relate to chemistry ("mixture," "liquid" and "pH") and physics ("velocity").



C

Term	Field	Area
Carbohydrate	Biology	14.52
Mixture (chemistry)	Chemistry	14.42
Protein	Biology	13.02
Velocity	Physics	12.08
Liquid	Chemistry	10.86
Enzyme	Biology	10.56
pH	Chemistry	9.95
X-ray	Physics	8.85
Density	Physics	8.72
Metabolism	Biology	7.10
Gravity	Physics	5.82
Fat	Biology	5.71
Cell (biology)	Biology	5.63
Homeostasis	Biology	5.29
Light spectrum	Physics	5.19
DNA	Biology	4.69
State of matter	Chemistry	4.63
Chemical elements	Chemistry	4.10
Ozone	Chemistry	3.82
Electrical insulator	Physics	3.63
Mass	Chemistry	3.49
Voltage	Physics	3.42
Genetic disorder	Biology	3.07
Gas	Chemistry	2.72
Pressure (chemistry)	Chemistry	2.29
Time (physics)	Physics	1.62
Volume (chemistry)	Chemistry	1.59
Electrical network	Physics	1.29
Radiation	Physics	0.59
Menstrual cycle	Biology	0.22

↑ Most dissimilar terms

Most similar terms ↓

Figure 1. LDA of search terms (A) by language (colors) and (B) by field (shapes). (C) Areas of the triangles between equivalent terms in different languages. LD1 and LD2: First and second linear discriminants, respectively.

Limitations and Concluding Remarks

The main limitation of this study derives from the measurement from a single point of access. Future studies could be conducted using pre-programmed searches from multiple servers to control for individual server influences (as performed by Scherr, Haim, & Arendt, 2019, for example).

Despite this limitation, this study provides a preliminary characterization of the quality of scientific information available to internet users in three languages. Our findings raise concern about digital inequalities in educational opportunities between students within the same country along lines of language proficiency, especially with respect to health and nutrition. The findings also suggest that disparities in the pedagogical quality of online content in the learner's language may contribute to second-level digital divides, especially among young learners. This would add another layer to the "digital inequality stack" of early exposure to computers, on top of known layers such as socio-economic status and gender (OECD, 2015; Robinson et al., 2020).

Further studies could explore the extent to which disparities on online content extend to additional languages and to other topics. Socio-scientific issues, such as COVID-19 and climate change, may be of special interest due to their importance for policymaking and individual decision-making.

Lastly, the findings can serve as a call to action for the scientific and educational communities and other institutions with respect to their public engagement. Our findings bolster Márquez & Porras' (2020) call to make scientific outreach initiatives more inclusive and multilingual. If we wish to help all learners achieve science literacy, we must act to mitigate the language divide.

References

Baram-Tsabari, A., & Yarden, A. (2005). Characterizing children's spontaneous interests in science and technology. *International Journal of Science Education*, *27*(7), 803–826. https://doi.org/10.1080/09500690500038389

Chen, W., & Wellman, B. (2004). The Global Digital Divide – Within and Between Countries. *IT & Society*, *1*(7), 39–45.

De Jesus, M., & Xiao, C. (2012). Predicting Internet use as a source of health information: A "language divide" among the Hispanic population in the United States. *Policy and Internet*, *4*(2). https://doi.org/10.1515/1944-2866.1178

Eberhard, D. M., Simons, G. F., & Fennig, C. D. (Eds.). (2020). *Ethnologue: Languages of the World* (23rd ed.). SIL International. Retrieved from http://www.ethnologue.com/

Ganayem, A. N. (2018). *HaInternet BaH̱evra Ha'Aravit BeYisra'el [The internet in Arab society in Israel]*. Petaẖ Tikva, Israel. Retrieved from https://www.isoc.org.il/research/policy-papers/the-internet-in-arab-society-in-israel

Guardiola-Wanden-Berghe, R., Gil-Pérez, J. D., Sanz-Valero, J., & Wanden-Berghe, C. (2011). Evaluating the quality of websites relating to diet and eating disorders. *Health Information & Libraries Journal*, *28*(4), 294–301. https://doi.org/10.1111/j.1471-1842.2011.00961.x

Hargittai, E., & Hsieh, Y. P. (2013). Digital Inequality. In W. H. Dutton (Ed.), *The Oxford handbook of Internet studies* (pp. 129–150). Oxford, England: Oxford University Press.

Israel Central Bureau of Statistics. (2013). *Selected Data from the 2011 Social Survey on Mastery of the Hebrew Language and Usage of Languages*. Jerusalem, Israel. Retrieved from https://www.cbs.gov.il/he/mediarelease/DocLib/2013/017/19_13_017b.pdf

Israel Central Bureau of Statistics and Israel National Authority for Measurement and Evaluation in Education. (2016). Preliminary Findings from the Israeli Survey of Adult Skills, 2014-2015 (Press Release). Retrieved November 24, 2018, from http://meyda.education.gov.il/files/Rama/PIAAC_Takzir.pdf

Israel Ministry of Science Technology and Space. (2017). Tfisot VeEmdot HaTsibur BeYisra'el BeNos'ey Madda Tekhnologia VeH̱alal [Attitudes and Perceptions of the Israeli Public toward Science, Technology and Space]. Jerusalem, Israel.

Lissitsa, S. (2015). Patterns of digital uses among Israeli Arabs – between citizenship in modern society and traditional cultural roots. *Asian Journal of Communication*, *25*(5), 447–464. https://doi.org/10.1080/01292986.2014.981555

Márquez, M. C., & Porras, A. M. (2020). Science Communication in Multiple Languages Is Critical to Its Effectiveness. *Frontiers in Communication*, *5*(May). https://doi.org/10.3389/fcomm.2020.00031

National Academies of Sciences Engineering and Medicine. (2016). *Science Literacy*. (C. E. Snow & K. A. Dibner, Eds.). Washington, D.C.: The National Academies Press. https://doi.org/10.17226/23595

National Science Board. (2020). Science and Technology: Public Attitudes, Knowledge, and Interest. In *Science and Engineering Indicators 2020*. Alexandria, VA: National Science Foundation. Retrieved from https://ncses.nsf.gov/pubs/nsb20207/

OECD. (2015). Inequalities in Digital Proficiency: Bridging the Divide. In *Students, Computers and Learning: Making the Connection* (pp. 123–143). Paris, France: OECD Publishing. https://doi.org/10.1787/9789264239555-8-en

Pearce, K. E., & Rice, R. E. (2014). The language divide-the persistence of english proficiency as a gateway to the internet: The cases of Armenia, Azerbaijan, and Georgia. *International Journal of Communication*, *8*(1), 2834–2859.

Robinson, L., Schulz, J., Blank, G., Ragnedda, M., Ono, H., Hogan, B., … Khilnani, A. (2020). Digital inequalities 2.0: Legacy inequalities in the information age. *First Monday*, *25*(7). https://doi.org/10.5210/fm.v25i7.10842

Savolainen, R. (2011). Judging the quality and credibility of information in Internet discussion forums. *Journal of the American Society for Information Science and Technology*, *62*(7), 1243–1256. https://doi.org/10.1002/asi.21546

Scherr, S., Haim, M., & Arendt, F. (2019). Equal access to online information? Google's suicide-prevention disparities may amplify a global digital divide. *New Media and Society*, *21*(3), 562–582. https://doi.org/10.1177/1461444818801010

Segev, E., & Sharon, A. J. (2016). Temporal patterns of scientif ic information-seeking on Google and Wikipedia. *Public Understanding of Science*. https://doi.org/10.1177/0963662516648565

Shahbazi, M., Farajpahlou, A., Osareh, F., & Rahimi, A. (2019). Development of a scale for data quality assessment in automated library systems. *Library & Information Science Research*, *41*(1), 78–84. https://doi.org/10.1016/j.lisr.2019.02.005

Warf, B., & Vincent, P. (2007). Multiple geographies of the Arab Internet. *Area*, *39*(1), 83–96. https://doi.org/10.1111/j.1475-4762.2007.00717.x

Warschauer, M. (2002). Reconceptualizing the Digital Divide. *First Monday*, *7*(7). https://doi.org/10.5210/fm.v7i7.967