*A novel data solution to analyse curriculum decolonisation – The case of Imperial College
London Masters in Public Health*

Robyn Price^{*a}, Mark Skopec^{*b}, Simon Mackenzie^a, Coco Nijhoff^a, Ruth Harrison^a, Gemma Seabrook^c and Matthew Harris^b

^a Library Services, Imperial College London, United Kingdom

^b School of Public Health, Faculty of Medicine, Imperial College London, United Kingdom

^c Information and Communication Technologies, Imperial College London, United Kingdom

Abstract

Analyses of reading lists by some UK Higher Education institutions in attempt to identify bias in curricula have found a prevalence of articles from the global north. However, previous studies have employed resource-intensive audit and data collection methods such as the authors or volunteers manually searching for and tagging individual reading list items by characteristic such as author country or place of publication. This can be prohibitive to repeating the study at different time periods or on large reading list data sets, which leads to a gap in evidence-based data to support and inform curriculum decolonisation. We describe a novel computational method applied to 568 articles, representing 3,166 authors from the Imperial College London Masters in Public Health (MPH) programme over two time periods (2017-18 and 2019-20). Using summary statistics, we found a shift in composite geographic distribution of reading lists sources across the two time periods studied and relate this to interventions to decolonise the

curriculum at Imperial. Our approach to applying a computational method to produce data as evidence in decolonisation toolkits is discussed.

Introduction

In recent years, staff and students at different Higher Education institutions (HEIs) in the UK have attempted to identify and dismantle legacies of colonisation embedded in institutional physical spaces, educational provisions and financial power structures (Carrell, 2019; *Decolonise UCL*, n.d.; *English Faculty Begins Decolonisation Discussion*, 2017; #RHODESMUSTFALL, n.d.; Sabaratnam, 2017). Projects including University College London's *Why is my curriculum white?* and *Diversifying and decolonising the medical curriculum* initiatives (Gishen & Lokugamage, n.d.; *Why Is My Curriculum White? - UCL - Dismantling the Master's House*, n.d.), the University of Cambridge's open letter to the Faculty of English (*DECOLONISING THE ENGLISH FACULTY*, n.d.), SOAS University of London's Learning and Teaching Toolkit for Programme and Module Convenors (Decolonising SOAS Working Group, 2018), Keele's University's Manifesto for decolonising the curriculum (*Keele Decolonising the Curriculum Network*, n.d.), University of Westminster discipline-specific suggested reading lists (University of Westminster, n.d.) and University of the Arts London's Decolonise the Arts Curriculum zines (*Decolonising the Arts Curriculum*, n.d.) indicate the diversity of responses to decolonise specifically the curriculum and suggest that each organisation can contribute unique responses to their individual needs. Whilst action has mostly been within individual HEIs, there is a growing cross-HEI community to support and inform practice; events such as 'Applying a Decolonial Lens to Research Structures, Norms and Practices in Higher Education Institutions' organised by SOAS (n.d.) and 'Decolonising the Curriculum – the library's role' organised by Goldsmiths University (*Decolonising the Curriculum - the Library's Role*, n.d.). The interconnections of

curriculum biases with the structural career barriers faced by black and minority ethnicity academics (Arday, 2020) and the lack of diversity in the academic library and publishing sectors (Charles, 2019) are important issues for sector-wide reflection. Despite these many examples of progress and support, there exists still obstacles such as hostility from within the academy either towards the concept of educational decolonisation or towards the people trying to install it (Liyanage, 2020) and only 24 out of 128 UK universities surveyed reported that they were committed to curriculum decolonisation (Batty, 2020).

There is considerable debate amongst decolonial scholars regarding the true meaning and mission of the decolonise education movement (Tuck & Yang, 2012). There is also opinion that HEIs reproduce colonial logic, and so any of our tools and methods cannot truly dismantle colonialism (Grewal, 2021). By function of deciding what is correct and what is taught, and conversely what is not correct and not taught, universities are not neutral spaces. However as stakeholders and in positions responsible for delivering education in an organisation we respond to the demand for ‘cognitive justice,’ represented by a community consisting of a plurality of knowledge, rather than an epistemic monoculture (Demeter, 2019). In many ways the global academic publishing economy has contributed to western monopolisation of access to and production of knowledge (Collyer, 2018; Demeter, 2019). 70% of the top 57 publishing houses are situated in the global north (Collyer, 2018). The spread of commercially owned journal publishing as a means of knowledge dissemination in some regions of Asia and South America has replaced previous traditions of local, institution- or state-owned not for profit publishing methods (Collyer, 2018). Furthermore, the monopolisation of publishing by commercial publishing houses has led to inflated subscription and publication costs, making participation for

scholars in the global south prohibitively expensive (Collyer, 2018). It has made penetrating the centralised global north market with research from the global south less feasible, hampering efforts to more widely disseminate articles from the global south (Collyer, 2018).

Bibliometric analyses of citation networks also demonstrate a gap in the consumption of global scientific research (Antonio Gálvez et al., 2000; Cash-Gibson et al., 2018; Keiser et al., 2004; Pan et al., 2012). A 2012 analysis found that North America and Europe receive 42.3% and 35.3% of the world's citations, respectively, compared to less than 5% by Africa, South America and Oceania combined (Pan et al., 2012). A 2019 analysis found that more than 75% of social science articles indexed in the Web of Science (WoS) database originated in either America or Western Europe, with the global south combined representing less than 10% (Demeter, 2019). Citation analyses further show that scholars of the global north are self-citing and inward-looking, while their colleagues in the global south primarily also cite research from 'core countries' (such as the UK and the US), rather than their own (Collyer, 2018). This unequal knowledge hierarchy is termed by historian Dipesh Chakrabaty as 'asymmetric ignorance'; western academics can afford to not cite non-western academics, without this affecting the perceived quality of their work, but non-western academics are not afforded this (Chakrabarty, 2000).

The gap in research consumption may be explained by a country-of-origin effect, studied extensively in consumer economics. Consumers evaluate products differently based on their country of origin and products from low-income countries (LICs) are generally rated less favourably due to this external cue (Adina et al., 2015; Bilkey & Nes, 1982; Srinivasan et al., 2004; Verlegh & Steenkamp, 1999). While country of origin effects have been extensively studied for physical products such as cars, wine or electronics (Verlegh & Steenkamp, 1999),

they may also extend to intellectual products, such as scientific research. For example, the country in which the research is produced and the institutional affiliation of the authors may be interpreted, consciously or unconsciously, by consumers of research as an indicator of the quality of the research (M. Harris, Marti, et al., 2017). Under controlled conditions, it has been demonstrated that research produced in high-income countries (HICs) is evaluated more favourably than the same research produced in LICs (M. Harris, Marti, et al., 2017), adjusted for other characteristics. It follows that a geographic bias in university curricula might be broadly reflective of this country of origin effect.

Empirical analysis of reading lists to understand geographic distribution of authors with the intent of revealing a bias (or lack of bias) is not yet widely used in practice. They could be an important tool for efforts to decolonise curricula (Schucan Bird & Pitman, 2019) in providing quantitative data as evidence in understanding whether or not certain geographic regions may be being unduly excluded, intentionally or not, from the university curriculum. Librarians at the London School of Economics found that 98.38% of all books and chapters of all course reading lists for the academic year 2019-20 were published by publishing houses based in global north countries, with 87.77% of this represented by the UK and US (Wilson, 2020). A 2019 analysis of two module reading lists (one social science, one STEM) at another UK university similarly revealed a skew that suggested that reading lists were dominated by white, male, Eurocentric authors (Schucan Bird & Pitman, 2019). Both studies were performed through manual review and coding of citations. The latter study concluded that “*Universities need to invest in research to develop and trial methods for interrogating diversity in module reading lists*” (Schucan Bird & Pitman, 2019). A study of gender bias in curriculum has produced a method of coding based on Application Programme Interfaces (APIs) (J. K. Harris et al., 2020) that whilst advancing on

resource-intensive manual methods deployed by previous studies (Phull et al., 2019), also call for *“Development of more sophisticated technological tools for automating the examination of syllabi might be a longer-term and more resource intensive solution”* to aide examination of gender bias in curricula, with which see a parallel with our efforts to examine geographic bias in curricula.

We propose a computational method that permits quasi-automated analysis of the geographical distribution of reading list authors. The method generates a quantitative indicator that facilitates time-specific and evidence-based interpretations of the data that can be used to supplement experience and theory-led decolonisation work. To clarify, we do not intend quantitative data to supplant the primacy of experience and theory in decolonisation, rather to support it. In this article, we describe the methods used to convert reading lists of the Imperial College London Masters in Public Health (MPH) programme over two time periods (2017-18 and 2019-20) into machine readable code from which bibliographic and author region data is retrieved from WoS and country socioeconomic status is retrieved from the World Bank. Using 1,200 articles we describe the shift in overall origin of reading list sources and relate this to interventions to decolonise the curriculum at Imperial College London during that period.

Methodology

We chose to analyse the MPH program as MH is a course co-Director, permitting a certain amount of ‘academic jurisdiction’ over the reading lists and because efforts to engage faculty in the MPH programme with decolonisation debates had been active since in 2018. Additionally, MPH programmes offer a varied subject matter drawing on geographically diverse research topics and therefore there was empirical interest to see if reading lists reflected that to some degree. Data was retrieved from two sources, 1) Leganto Reading Lists (*Leganto Reading*

Lists, n.d.) (henceforth “Leganto”), a digital reading list management system and 2) the Web of Science Core Collection (WoS) (‘Web of Science Core Collection’, n.d.).

Stage One – retrieve items from reading list management system

We searched Leganto using a reporting tool (Oracle Business Intelligence Enterprise Edition) for “Academic Department Description” = “School of Medicine” AND “Course Year” = ‘2018’ or ‘2019’ or ‘2020’. This search retrieved data associated with all courses in the School of Medicine faculty. Limiting to the selected module reading list (Masters of Public Health) is performed in Stage 2. The search was in the form of a SQL request that retrieved relevant bibliographic and course fields associated with each found reading list item with a DOI. Results were exported in Excel .xlsx format (Leganto_wos_doi.xlsx) to a specified folder.

Table 1 SQL request to retrieve bibliographic and course data

SELECT

0 s_0,

"Course Reserves"."Citation Metadata Details"."Author" s_1,

"Course Reserves"."Citation Metadata Details"."DOI" s_2,

"Course Reserves"."Citation Metadata Details"."Journal Title" s_3,

"Course Reserves"."Citation Metadata Details"."Publication Date" s_4,

"Course Reserves"."Citation Metadata Details"."Title" s_5,

"Course Reserves"."Course"."Academic Department Description" s_6,

"Course Reserves"."Course"."Course Code" s_7,

"Course Reserves"."Course"."Course Instructor Primary Identifier" s_8,

"Course Reserves"."Course"."Course Status" s_9,

"Course Reserves"."Course"."Course Year" s_10,

"Course Reserves"."Course"."Searchable Id 1" s_11,

"Course Reserves"."Reading List Citation"."Citation Type" s_12,

```

"Course Reserves"."Reading List"."Reading List Code" s_13,
"Course Reserves"."Reading List"."Reading List Name" s_14
FROM "Course Reserves"
WHERE
(("Citation Metadata Details"."DOI" IS NOT NULL) AND ("Course"."Academic
Department Description" = 'School of Medicine') AND ("Course"."Course Year" IN
('2018', '2019', '2020')))
ORDER BY 1, 6 ASC NULLS FIRST, 3 ASC NULLS FIRST, 2 ASC NULLS FIRST, 5
ASC NULLS FIRST, 8 ASC NULLS FIRST, 14 ASC NULLS FIRST, 15 ASC NULLS
FIRST, 4 ASC NULLS FIRST, 9 ASC NULLS FIRST, 12 ASC NULLS FIRST, 7 ASC
NULLS FIRST, 10 ASC NULLS FIRST, 11 ASC NULLS FIRST, 13 ASC NULLS FIRST
FETCH FIRST 500001 ROWS ONLY

```

Stage 2 – retrieve author-specific bibliographic data

A PHP script imports the data from each file in this folder and makes a separate curl HTTP request to the WoS database for each reading list item via a GET command to the WoS API Expanded. An example of the GET command format in this retrieval:

```

https://wos-
api.clarivate.com/api/wos/?databaseId=WOK&count=1&firstRecord=1&usrQuery=DO=
%2210.1177/1075547010389818%22.

```

DOIs must be wrapped by URL encoded quotation marks (%22) to avoid null-returns from the WoS for DOIs containing parenthesis.

This command automates the sending of all identified reading list items to the WoS to retrieve item-level metadata. The relevant returned fields for this analysis are “Author Sequence”, “Author Address”, “Author City” and “Author Country”. These fields were extracted from the XML of matched responses from the WoS response and merged with the Leganto

extract of Stage 1, then exported to a CSV file format. The full PHP script is available in Supplementary Data “PHP input.txt”.

Stage 3 – retrieve country socioeconomic indicator

The results of the automated searching and matching in Stages 1 and 2 resulted in CSV files for each of the course modules and years. We then used data generated by the World Bank which ranks countries based on the Gross National Income (GNI) per capita (Atlas Method)(*GNI per Capita, Atlas Method (Current US\$) / Data*, n.d.), to compose numeric value to citations based on the distribution of authors’ affiliations.

Stage 4 – transform data

We matched World Bank regions to WoS author affiliation regions. 40 data points were removed as they were geographic regions that came from the WoS author country field that either did not refer to a country for which World Bank data was available (such as “Arab World”, “Middle East and Northern Africa” or “Latin America & Caribbean”), or the World Bank classification returns no data (Andorra, Gibraltar or Monaco, among others). This left a remainder of 203 countries, ranked by income. For 184 countries, income data from 2018 was used. For 19 countries, 2018 data was not available, and the most recent available data was used instead. A list of removed countries, as well as a list of which countries listed most recent non-2018 data, can be found in Supplementary Data “World Bank_GNI per capita_2020_.xls”

The CSV files generated in Stages 1 and 2 were also subject to data cleansing, as not all the information retrieved was necessary for our analysis. Unique module identifiers (also known as ‘module codes’) and module titles were retained to allow for easier sorting and subsequent

comparisons between modules. The DOI and citation title were retained in order to conduct periodic spot-checks to ensure that the methods outlined in Stages 1 and 2 had been successful.

Author institutional affiliation, city and country were obtained through the WoS command, but we only used the country-level data of the authors for our analysis. The WoS output provided country data for each author listed on a citation, however, in the CSV file, this information, for all authors of a citation, was provided in the same cell. Prior to analysis, we therefore separated this data out into individual cells, adhering to author order as specified in the citation. Where authors possessed multiple affiliations, only the first listed affiliation was included in the analysis. The country of each author's affiliation was then replaced with its corresponding inverse rank according to the GNI/capita ranking. For instance, "Australia" was replaced with its corresponding rank, 188. Where WoS disambiguates author territories in a different way to the World Bank, but both were identifiable, we compressed the regions to match the World Bank schema, e.g. WoS disambiguates "England," "Wales," "Scotland" and "Northern Ireland" separately, whereas World Bank includes all under "United Kingdom" and all were assigned the same rank, 179. The percentile ranking was then used to calculate the Citation Source Index as described below.

Citation Source Index

The institutional affiliation data generated through Leganto and WoS (Stages 1 and 2), as well as the GNI/capita data (Stage 3), were used to calculate a Citation Source Index (CSI). This CSI represented a weighted average of the World Bank rankings for GNI/capita of the countries where the authors in that citation were from. A formula for the CSI can be seen in Equation 1. As there were 203 countries included in the World Bank ranking, the number was represented in the denominator of the equation.

$$\text{Citation Source Index (CSI)} = \frac{(\text{GNI rank}_{\text{Author 1}} + \text{GNI rank}_{\text{Author 2}} + \dots + \text{GNI rank}_{\text{Author } \mu})}{(203 * \sum \text{Authors})}$$

Equation 1. Citation Source Index (CSI). The CSI provides a weighted score for each individual citation, based on the citation's authors' institutional affiliations. The closer the CSI is to 1, the more high-income countries are represented on a given citation.

Based on this calculation, a CSI could be any value between 0.0049 (for a citation with authors exclusively from Somalia, ranked at number 1 on the GNI/capita list) and 1 (for a citation with authors exclusively from Liechtenstein, ranked number 203 on the GNI/capita list). A CSI closer to 1 would represent authorship primarily affiliated with institutions in HICs, whereas a CSI closer to 0.0049 would represent authorship primarily affiliated with institutions in LICs. Calculating a CSI for each module on the course, as well as calculating an aggregate CSI for both the 2017-18 and 2019-20 course, permits easily visualisation to describe any changes in the geographical distributions of the reading lists.

Findings

Data were analysed at the course and module level. For the purpose of this paper, modules within the Masters of Public Health have been de-identified. Box and whiskers plots were generated to allow for simple visualisation of the median CSI, the interquartile range (IQR), as well as outliers on the course or module. Below, we discuss the 2017-18 and 2019-20 results individually. We also evaluate outliers and discuss potential reasons for the discrepant scores and compare the changes in CSI for specific modules.

‘Missing’ data

Not all reading list items from the 2017-18 and 2019-20 MPH could be represented in the analysis because we only extracted journal article data from the WoS. For example, the 2017-18 dataset consisted of 660 total citations and 354 (53.63%) of these citations were journal articles and were therefore viable for analysis. The 2019-20 dataset consisted of 553 total citations, however three modules did not have their reading lists published on Leganto, so data for these modules could not be retrieved, resulting in a smaller sample size compared to 2017-18. 314 (56.78%) of the 2019-20 citations were journal articles and were therefore viable for analysis.

For the 2017-18 dataset, of the 354 journal articles 37 (10.45%) articles did not generate data through WoS. An additional 3 articles (0.84%) were excluded as the authors did not have affiliations listed on WoS, or because the country of affiliation was not listed in the World Bank's ranking. This occurred, for instance, when the author's affiliation on WoS was listed as "Taiwan" or "Occupied Palestinian Territories." 314 articles (or 89.54% of the viable articles containing DOIs) were included in the final analysis.

For the 2019-20 dataset, of the 323 journal articles 65 (20.70%) articles did not generate data through WoS. An additional 4 articles (1.274%) were excluded from analysis because authors did not have an affiliation listed on WoS, or because the country of affiliation was not listed in the World Bank's ranking. 254 articles (or 80.89% of the viable articles containing

DOIs) were included in the final analysis. Figure 1 depicts a flowchart of this breakdown.

2017-18 Data

2019-20 Data

Figure 1. Flowchart of missing data. 46.36% and 42.53% of total citations were excluded for 2017-18 and 2019-20, respectively, as these citations did not possess a DOI. 88.70% and 78.64% of viable citations for 2017-18 and 2019-20, respectively, were included in the final analysis.

Manual spot-checks of 10% of citations for each year were undertaken to ensure that no errors existed in the country affiliation data. Citations were picked using a random number generator created in Excel. For 2017-18, one citation (0.318% of the total included for that year) only retrieved information for one author, despite there being five authors on the paper. As all authors were affiliated with institutions in the UK, this did not change the CSI. For 2019-20, one citation (0.394% of the total included for that year) was attributed to an author affiliated with an institution in Kenya, but spot-checking revealed that they were in fact affiliated with an institution in the UK. This was an error attributed to WoS indexing which has since been corrected by WoS and corrected in our dataset. Aside from these errors, no additional errors were found. Tables 1 and 2 show a module-by-module breakdown of total viable citations for the study (e.g. they have been a journal article, have a DOI), percentage of viable citations not found

by the WoS and the total authors represented. Table 1 and Table 2 show a complete module-by-module breakdown of all citations for each module, including citation type.

Course code	Total citations	Citations missing (%)	Total authors
MA2017-18	75	9 (12.00)	291
MB2017-18	19	3 (15.79)	35
MC2017-18	5	0	9
MD2017-18	53	6 (11.32)	278
ME2017-18	27	0	153
MF2017-18	26	1 (3.846)	58
MG2017-18	29	7 (24.14%)	68
MH2017-18	22	1 (4.545)	198
MI2017-18	18	1 (5.555)	76
MJ2017-18	15	5 (33.33)	31
MK2017-18	1	0	1
ML2017-18	2	0	23
MM2017-18	9	0	314
MN2017-18	41	3 (7.317)	263
MO2017-18	12	1 (8.333)	26
MP2017-18	-	-	-
TOTAL	354	37	1824

Table 1. 2017-18 Summary. This table summarises the viable citations that were included in the analysis. 37 citations were missing, and 3 citations were excluded as there was no author information available. 314 citations, comprising 1566 author were ultimately analysed. A more detailed summary, which includes the types of citations on each individual module, can be found in Supplementary Data.

Course code	Total citations	Citations missing (%)	Total authors
MA2019-20	81	17 (20.99)	277
MB2019-20	1	0	2
MC2019-20	-	-	-
MD2019-20	75	10 (13.33)	374
ME2019-20	44	18 (40.91)	139
MF2019-20	27	2 (7.407)	1354
MG2019-20	9	0	32
MH2019-20	3	1 (33.33)	105
MI2019-20	16	4 (25)	73
MJ2019-20	-	-	-
MK2019-20	2	2 (100)	0
ML2019-20	2	0	10
MM2019-20	8	0	30
MN2019-20	40	3 (7.5)	233
MO2019-20	15	8 (53.33)	22
MP2019-20	-	-	-
TOTAL	323	65	2651

Table 2. 2019-20 Summary. This table summarises the viable citations that were included in the analysis. 65 citations were missing, and 4 citations were excluded as there was no author information available. 254 citations, comprising 1342 authors were analysed. A more detailed summary, which includes the types of citations on each individual module, can be found in Supplementary Data.

Results

314 citations were analysed for the 2017-18 reading lists, comprising 1566 authors. The median CSI for 2017/18 was 0.8818 (IQR 0.8818 – 0.9498). The mean CSI was 0.8837, the maximum was 0.9901, and the minimum was 0.4712. 254 citations were analysed for the 2019-20 reading lists, comprising 1342 authors. The median CSI for 2019/20 was 0.8818 (IQR 0.8818 – 0.9557). The mean CSI was 0.8803, the maximum was 0.9901, and the minimum was 0.4778. We also compared the composition of several modules between 2017-18 and 2019-20, to establish whether there was a shift in the geographic distribution of selected articles over this time period. Figure 2 shows a comparison of CSI scores for the course overall, as well as for the individual modules that were analysed.

Figure 2. Comparison of CSI scores from 2017-18 and 2019-20. The above graph depicts a comparison of aggregate CSI scores for the 2017-18 and 2019-20 MPH course ("ALL"), as well as a comparison of individual modules. MC2019-20, MJ2019-20, MK2019-20, and MP2017-18 and 2019-2020 did not contain citations for analysis. This is discussed further in the Limitations section.

A skew toward research produced in countries in the global north is observed. However, the course median CSI did not change between 2017-18 and 2019-20 (0.8817), though mean CSI was slightly lower in 2019-20 (from 0.8837 to 8.8803, a decrease by 0.0034 points). There was also very little change in CSI at the module level. Both median and mean scores remained were similar over the two-year time interval. Median CSI decreased in four modules, ME2017-18/2019-20 (N=27 and 26 citations, respectively), ML2017-18/2019-20 (N=2 and 2 citations,

respectively), MM2017-18/2019-20 (N=9 and 8 citations, respectively) and MN2017-18/2019-20 (N=38 and 37 citations, respectively). The median CSI was lowered by 0.0167, 0.0026, 0.083 and 0.0099 points, respectively. Median CSI increased for , MA2017-18/2019-20 and MH2017-18/2019-20 by 0.0739 and 0.0114, respectively. For the remaining three modules, median CSI remained the same.

Mean CSI decreased in MB2017-18/2019-20 (N=16 and 1, respectively), ME2017-18/2019-20 (N=27 and 26, respectively), MI2017-18/2019-20 (N=17 and 12, respectively), , ML2017-18/2019-20 (N=2 and 2, respectively) MM2017-18/2019-20 (N=9 and 8, respectively), and MO2017-18/2019-20 (N=11 and 7, respectively). Mean CSI decreased by 0.0034 overall, and by 0.0035 (MB2017-18/2019-20), 0.0005 (ME2017-18/2019-20), 0.0059 (MI2017-18/2019-20), 0.083 (ML2017-18/2019-20), 0.0585 (MM2017-18/2019-20) and 0.0209 (MO2017-18/2019-20) for the modules. . On the other hand, mean CSI increased for MA2017-18/2019-20 (0.0009), MD2017-18/2019-20 (0.0044), MF2017-18/2019-20 (0.0096), MG2017-18/2019-20 (0.0024) MH2017-18/2019-20 (0.0047), and MN2017-18/2019-20 (0.0135).

Overall, the modules continued to display a skew toward research produced by authors affiliated with institutions in HICs. However, though the CSI scores indicate both median and mean scores centred nearer to 1, the long ‘tails’ in Figure 2 indicate that most modules do still include a variety of citations, including some with authors affiliated with institutions in LICs. Particularly in the example of MI2017-18/2019-20, the ‘addition’ of a tail between the two points in time may seem to indicate a greater diversity of readings, and more inclusion of research with authors from LICs. Indeed, the minimum score for this module shifted from 0.8818 in 2017-18 to 0.775 in 2019-20, following the addition of a paper featuring a collaboration between authors from the US, the UK, Ireland and South Africa. Yet, the opposite appears to be

true for MF2017-18/2019-20, as there are very narrow tails, and there appears to have been no addition of citations featuring LIC authors between the two time points.

ME2017-18 contains the citation with the lowest score for 2017-18. This citation (*India's Conditional Cash Transfer Programme (the JSY) to Promote Institutional Birth: Is There an Association between Institutional Birth Proportion and Maternal Mortality?*, DOI: 10.1371/journal.pone.0067452) features a collaboration between two authors based at Gardi Medical College in Ujjain, India, and one author based at Karolinska Institutet in Umea, Sweden. However, this citation is no longer included in the 2019-20 reading list. MN2017-18/2019-20 contains the citation with the lowest score for 2019-20 (0.477833). This citation (*Barriers for Adopting Electronic Health Records (EHRs) by Physicians*, DOI: 10.5455/aim.2013.21.129-134) features a co-authorship between two academics at Isfahan University of Medical Sciences in Isfahan, Iran.

Discussion

We found that for both time periods, citations on the reading list were authored primarily by institutions from the global north, and there was marginal change in the geographic distribution of citations between the two time points. The methodology was successful at providing a quantitative analysis of the geographic origin of citations that are indexed in WoS. Data was missing from the citations that appeared in peer-reviewed journals (10% of 2017-18 of citations with a DOI and 20% of 2019-20 citations with a DOI), however spot-checks demonstrated a very high level of accuracy in terms of author affiliation attributed in WoS. Accounting for this missing data, the descriptive analysis was performed for over 50% of the entirety of all items on the module reading lists over the two time periods.

As far as we are aware, the methods detailed in this paper are the first to connect a reading list management programme, a bibliographic database and World Bank data in a semi-automated method and therefore could be a novel addition to decolonise education toolkits, both in terms of speed of data collection and analysis, as well as the intersecting territorial income schema with curriculum as an indicator of diversity. Automation allows descriptive analysis on a large scale and the CSI permits comparison of trends in geographic origin of the readings provided in a course over time periods. The lack of significant change in CSI at course and module level is unsurprising given efforts to engage faculty responsible for this course with debates surrounding decolonisation began in earnest only at the end of 2018. It also indicates that literatures included in postgraduate courses are relatively static over time and a concerted effort is required to effect significant shift in reading list composition. Nevertheless, reading list analysis through data visualisation is useful to explore trends and stimulate debate.

In addition to understanding the contemporary legacies of scholars and institutions as perpetrators or victim of colonialism, we must also investigate the implicit power structures in how knowledge is produced, consumed, distributed and assessed, and the bearing this might have on perpetuating inequality. Particularly as it relates to the consumption of knowledge, the CSI we calculate here could offer an important departure point. The results of the MPH analysis indicate a skew toward research produced by researchers affiliated with institutions predominantly in HICs. Our findings are broadly reflective of the global research production landscape, which itself is heavily skewed toward the global north, as demonstrated by others (Cash-Gibson et al., 2018; Pan et al., 2012). However, we cannot state for certain whether the skew in the reading lists is because of that broader pattern. The descriptive nature of our analysis does not allow us to determine whether articles from the global south that were added to the reading list in 2019-20,

have been thoughtfully incorporated into the taught curriculum, or if they have been included, consciously or not, in anticipation of the CSI measurement in this project. Our quantitative method cannot meaningfully answer questions of motivation and therefore it cannot determine changes in attitudes. Whilst a shift in the CSI of a course over time from a value closer to 1 to a value closer to 0.0049 would suggest that the geographic distribution of a reading list's authors included a greater proportion of authors affiliated with institutions in LICs, the CSI is merely one way to describe a reading list. It is not intended to be a score of the reading list. Rather, it is intended to provide a baseline to assess the geographic composition of reading lists and a method to assess trends and changes over time. The CSI can give an indication of geographic representation that can support, but should not lead, qualitative and theoretical interpretation.

The method would permit the analysis of different course reading lists quickly and efficiently. However, we believe that the CSI is only of real value if it can be meaningfully interpreted by course leaders or students. They alone have the mix of subject matter expertise to decide if a) there is need or desire for a geographic based 'liberation lens' (Charles, 2019) to be applied to this course; b) if the material is appropriate for this method and its limitations and then finally c) to decide if the results are meaningful in the context of the subject matter, the composition of the taught student body, location of HEI and more. We are using the results of the single course analysis to begin conversations across Imperial College London to find groups interested to apply an empirical perspective to their curriculum and equipped to interpret the results in a meaningful way.

Other initiatives towards decolonisation at Imperial College London include a half-day faculty workshop, held in collaboration with the Educational Development Unit (EDU) on a biannual basis, which aims to provide a space for faculty and staff to have facilitated discussions

surrounding this topic. Not coincidentally, this very research project arose as a result of several members of the library staff (including CN) attending one of these workshops facilitated by MH and MS. We have designed a bespoke Implicit Association Test (IAT), which measures respondents' unconscious associations between "Good and Bad Research" and "High- and Low-income countries." To date, 201 members of the Imperial College London community have taken the test. The median D-Score (the measure of unconscious bias) of 0.527 (95% CI -0.249; 1.302), suggests a "Moderate" unconscious association between HICs and "Good research" among participants. A detailed discussion of the IAT methods we used have been documented elsewhere (M. Harris, Macinko, et al., 2017). Information Literacy teaching provision from the library is underpinned by the A New Curriculum for Information Literacy (ANCIL) framework (McConnell, 2011) and has been reviewed to ensure delivery of the Critical Analysis competency to recognise bias in authors and intellectual movements. A library-specific Decolonise the Academy event has been held to promote reflection and support for decolonisation action. College networks and bodies including Imperial as One and the Equality, Diversity and Inclusion Centre lead institutional reflection and change. None of these actions in isolation are definitive solutions to challenge bias in the institutional structure, but together they inform and equip staff and student to tackle the issue.

Limitations

Identification of articles present in the reading list is dependent on the course leaders having correctly registered the items by DOI in the Leganto reading list system, this is manual system and thus open to human error. This was the case for MC2019-20 data set. Any items entered with an incorrect or missing DOI will not return data unless manually corrected. Any reading list items not captured by course leaders in Leganto, e.g. items communicated to students outside of

the official course reading list system, were not retrievable. Items that do not have a DOI cannot return data. This excluded output types other than journal articles from the analysis. For the reading lists concerned, the commonly excluded output types included books, book chapters, grey literature such as presentation files, health policy reports and websites (see Figure 1). For this reason, this method would not be appropriate for courses with reading lists with a majority non-journal article outputs. In our study, this was the case for MP2017-18/2019-20, which relied only on textbooks, and was therefore not included in the analysis. However, as shown in Figure 1, more than 50% of all readings for both 2017-18 and 2019-20 were still viable for analysis. We are also aware of the shortcomings of using paid-for, close access data sources (Leganto and WoS) to evaluate decolonisation and would welcome sources that are free to access, modify and redistribute.

We are particularly conscious of the limitation of using the WoS database. A crucial data field returned in Stage 2, the Author Sequence (“addr_no”) permits the authors to be identified by their numerical positioning in the manuscript author list. This data field is not regularly supplied by WoS for items published before 2008. This affected 91 items for all reading list years studied in this analysis. A further 20 records returned null results because the journals they were published in are not included in WoS’ journal coverage. Although broadly comprehensive of medical research, WoS is still a selective source with coverage limited by an editorial selection process (criticism of WoS as a source is addressed in the Introduction). An alternative database that provides the extensive bibliographic data in a similar format to WoS, but with more inclusive or broader indexing and less prohibitive access costs would be welcome in future adaptations of the method.

Finally, a potential limitation of the study concept study is our definition of ‘geographic bias’ as limited to only to the assigned country of authors as declared on the published manuscript. This does not take into account any other demographic characteristic that could describe the author, such as race or ethnicity. Similarly, we equate an author’s institutional affiliation to be synonymous with their own nationality. We assume that an author based at a UK institution is British, and therefore their assigned rank is 179. This is of course not always the case. An author at a UK institution could originate from an LIC and could thus contribute a different perspective than their co-authors from the UK, but our methodology is not able to account for this. It is also not able to investigate the content of the papers we analysed. So, even though a study may have been authored by researchers in HICs, it could contain content that is valuable to a decolonial perspective. Application of natural language processing tools for academic outputs could potentially provide semantic analysis (Atanassova et al., 2019) to infer sentiment. The application of gender assessment methods (Sumner, 2018) could also be combined with our method to examine intersections of geographic and gender bias.

Conclusions

Our findings for the Imperial MPH reading list indicate a notable skew towards authors from HICs with a marginal reduction from the 2017-18 course year to the 2019-20 course. In producing these results we have also developed a method that permits the bulk of the labour preparing and extracting data for a reading list analysis to transfer from a human to a computer. This is important for empirical reading list analyses to become more accessible tools in curriculum decolonisation projects. We stress that any replication of this method should be considered by a group empowered to meaningfully interpret results for a course and as part of wider decolonisation reflections. A ‘high’ score in CSI metric is not necessarily indicative of a

biased curriculum, just as a ‘low’ score should not be taken as unassailable evidence of a diverse, inclusive, or ‘decolonised’ curriculum. Complex and often interconnected issues of diversity and legacies of colonialism in HEIs and curricula, as well as imbalances in knowledge production and access require deep and multimethod investigation. Data can provide justification for change and may be particularly strong incentive to encourage institutional investment of resources into decolonisation projects but should serve only as one form of evidence in a wider context, and cannot answer the question of what a reading list *should* look like.

References

- Adina, C., Gabriela, C., & Roxana-Denisa, S. (2015). Country-of-Origin Effects on Perceived Brand Positioning. *Procedia Economics and Finance*, 23, 422–427.
[https://doi.org/10.1016/S2212-5671\(15\)00383-4](https://doi.org/10.1016/S2212-5671(15)00383-4)
- Antonio Gálvez, Mercedes Maqueda, Manuel Martínez-Bueno, & Eva Valdivia. (2000). Scientific Publication Trends and the Developing World: What can the volume and authorship of scientific articles tell us about scientific progress in various regions? *American Scientist*, 88(6), 526–533. <http://www.jstor.org/stable/27858122>
- Arday, J. (2020). Fighting the tide: Understanding the difficulties facing Black, Asian and Minority Ethnic (BAME) Doctoral Students’ pursuing a career in Academia. *Educational Philosophy and Theory*, 1–8. <https://doi.org/10.1080/00131857.2020.1777640>
- Atanassova, I., Bertin, M., & Mayr, P. (2019). Editorial: Mining Scientific Papers: NLP-enhanced Bibliometrics. *Frontiers in Research Metrics and Analytics*, 4, 2.
<https://doi.org/10.3389/frma.2019.00002>

Batty, D. (2020, June 11). Only a fifth of UK universities say they are ‘decolonising’ curriculum.

The Guardian. <https://www.theguardian.com/us-news/2020/jun/11/only-fifth-of-uk-universities-have-said-they-will-decolonise-curriculum>

Bilkey, W. J., & Nes, E. (1982). Country-of-Origin Effects on Product Evaluations. *Journal of International Business Studies*, 13(1), 89–100.

<https://doi.org/10.1057/palgrave.jibs.8490539>

Carrell, S. (2019, August 23). Glasgow University to pay £20m in slave trade reparations. *The Guardian*. <https://www.theguardian.com/uk-news/2019/aug/23/glasgow-university-slave-trade-reparations>

Cash-Gibson, L., Rojas-Gualdrón, D. F., Pericàs, J. M., & Benach, J. (2018). Inequalities in global health inequalities research: A 50-year bibliometric analysis (1966-2015). *PLOS ONE*, 13(1), e0191901. <https://doi.org/10.1371/journal.pone.0191901>

Chakrabarty, D. (2000). *Provincializing Europe: Postcolonial thought and historical difference*. Princeton University Press.

Charles, E. (2019). Decolonizing the curriculum. *Insights*, 32(1), 24.

<https://doi.org/10.1629/uksg.475>

Collyer, F. M. (2018). Global patterns in the publishing of academic knowledge: Global North, global South. *Current Sociology*, 66(1), 56–73.

<https://doi.org/10.1177/0011392116680020>

Decolonise UCL. (n.d.). Students’ Union UCL. Retrieved 9 November 2018, from

<http://studentsunionucl.org/node/115789>

Decolonising SOAS Working Group. (2018). *Decolonising SOAS Learning and Teaching Toolkit for Programme and Module Convenors*.

<https://blogs.soas.ac.uk/decolonisingsoas/learning-teaching/toolkit-for-programme-and-module-convenors/>

Decolonising the Arts Curriculum: Perspectives on Higher Education. (n.d.). Retrieved 29 June 2020, from <https://decolonisingtheartscurriculum.myblog.arts.ac.uk/>

Decolonising the Curriculum—The library's role. (n.d.). Decolonising the Curriculum - the Library's Role. Retrieved 20 February 2020, from <https://decolonisethelibrary.wordpress.com/>

DECOLONISING THE ENGLISH FACULTY: AN OPEN LETTER. (n.d.). Google Docs. Retrieved 12 November 2018, from https://docs.google.com/document/d/1Qji9ojNzumOeKboLLBBWs5fxJfEPVAw4JDNtdz2yAtU/edit?usp=sharing&usp=embed_facebook

Demeter, M. (2019). The World-Systemic Dynamics of Knowledge Production: The Distribution of Transnational Academic Capital in the Social Sciences. *Journal of World-Systems Research*, 25(1), 111–144. <https://doi.org/10.5195/jwsr.2019.887>

English Faculty begins decolonisation discussion. (2017, October 25). The Cambridge Student. <https://www.tcs.cam.ac.uk/news/0037798-english-faculty-begins-decolonisation-discussion.html>

Gishen, F., & Lokugamage, A. (n.d.). *Diversifying and decolonising the medical curriculum—The BMJ*. Retrieved 14 March 2019, from <https://blogs.bmj.com/bmj/2018/10/18/diversifying-and-decolonising-the-medical-curriculum/>

GNI per capita, Atlas method (current US\$) | Data. (n.d.). Retrieved 30 January 2019, from <https://data.worldbank.org/indicator/NY.GNP.PCAP.CD>

Grewal, M. (2021, March 9). To ‘decolonise’ education, we need to ask different questions.

Times Higher Education. <https://www.timeshighereducation.com/opinion/decolonise-education-we-need-ask-different-questions>

Harris, J. K., Croston, M. A., Hutti, E. T., & Eyler, A. A. (2020). Diversify the syllabi:

Underrepresentation of female authors in college course readings. *PLOS ONE*, 15(10), e0239012. <https://doi.org/10.1371/journal.pone.0239012>

Harris, M., Macinko, J., Jimenez, G., & Mullachery, P. (2017). Measuring the bias against low-income country research: An Implicit Association Test. *Globalization and Health*, 13(1).

<https://doi.org/10.1186/s12992-017-0304-y>

Harris, M., Marti, J., Watt, H., Bhatti, Y., Macinko, J., & Darzi, A. W. (2017). Explicit Bias

Toward High-Income-Country Research: A Randomized, Blinded, Crossover Experiment Of English Clinicians. *Health Affairs*, 36(11), 1997–2004.

<https://doi.org/10.1377/hlthaff.2017.0773>

Keele decolonising the curriculum network. (n.d.). Keele University. Retrieved 29 June 2020, from

<https://www.keele.ac.uk/equalitydiversity/equalityawards/raceequalitycharter/keeledecolonisingthecurriculumnetwork/#keele-manifesto-for-decolonising-the-curriculum>

Keiser, J., Utzinger, J., Tanner, M., & Singer, B. H. (2004). Representation of authors and editors from countries with different human development indexes in the leading literature on

tropical medicine: Survey of current evidence. *BMJ*, 328(7450), 1229–1232.

<https://doi.org/10.1136/bmj.38069.518137.F6>

Leganto Reading Lists. (n.d.). Imperial College London. Retrieved 9 September 2020, from

<http://www.imperial.ac.uk/admin-services/library/learning-support/reading-lists/>

- Liyanage, M. (2020). *Miseducation: Decolonising curricula, culture and pedagogy in UK universities*. Higher Education Policy Institute. <https://www.hepi.ac.uk/2020/07/23/new-report-calls-for-decolonisation-of-universities-in-order-address-the-silent-crisis-in-universities/>
- McConnell, C. (2011, August 25). Use ANCIL in your institution. *A New Curriculum for Information Literacy*. <https://newcurriculum.wordpress.com/using-ancil/>
- Pan, R. K., Kaski, K., & Fortunato, S. (2012). World citation and collaboration networks: Uncovering the role of geography in science. *Scientific Reports*, 2(1). <https://doi.org/10.1038/srep00902>
- Phull, K., Ciflikli, G., & Meibauer, G. (2019). Gender and bias in the International Relations curriculum: Insights from reading lists. *European Journal of International Relations*, 25(2), 383–407. <https://doi.org/10.1177/1354066118791690>
- #RHODESMUSTFALL. (n.d.). #RHODESMUSTFALL. Retrieved 12 December 2018, from <https://rmfoxford.wordpress.com/>
- Sabaratnam, M. (2017, January 18). Decolonising the curriculum: What's all the fuss about? *SOAS Blog*. <https://www.soas.ac.uk/blogs/study/decolonising-curriculum-whats-the-fuss/>
- Schucan Bird, K., & Pitman, L. (2019). How diverse is your reading list? Exploring issues of representation and decolonisation in the UK. *Higher Education*. <https://doi.org/10.1007/s10734-019-00446-9>
- SOAS. (n.d.). *Applying a Decolonial Lens to Research Development Practices – SOAS Blog*. Retrieved 1 May 2020, from <https://www.soas.ac.uk/blogs/study/applying-a-decolonial-lens-to-research-development-practices/>

- Srinivasan, N., Jain, S. C., & Sikand, K. (2004). An experimental study of two dimensions of country-of-origin (manufacturing country and branding country) using intrinsic and extrinsic cues. *International Business Review*, 13(1), 65–82.
<https://doi.org/10.1016/j.ibusrev.2003.05.004>
- Sumner, J. L. (2018). The Gender Balance Assessment Tool (GBAT): A Web-Based Tool for Estimating Gender Balance in Syllabi and Bibliographies. *PS: Political Science & Politics*, 51(2), 396–400. <https://doi.org/10.1017/S1049096517002074>
- Tuck, E., & Yang, K. W. (2012). Decolonization is not a metaphor. *Decolonization: Indigeneity, Education & Society*, 1(1), 1–40. <https://www.latrobe.edu.au/staff-profiles/data/docs/fjcollins.pdf>
- University of Westminster. (n.d.). *Decolonising the curriculum toolkit*. Retrieved 1 January 2020, from <http://cti.westminster.ac.uk/decolonising-the-curriculum-toolkit/>
- Verlegh, P. W. J., & Steenkamp, J.-B. E. M. (1999). A review and meta-analysis of country-of-origin research. *Journal of Economic Psychology*, 20(5), 521–546.
[https://doi.org/10.1016/S0167-4870\(99\)00023-9](https://doi.org/10.1016/S0167-4870(99)00023-9)
- Web of Science Core Collection. (n.d.). *Web of Science Group*. Retrieved 9 October 2020, from <https://clarivate.com/webofsciencegroup/solutions/web-of-science-core-collection/>
- Why is my curriculum white? - UCL - Dismantling the Master's House*. (n.d.). Retrieved 1 May 2020, from <http://www.dtmh.ucl.ac.uk/videos/curriculum-white/>
- Wilson, K. (2020, January 24). *Decolonising LSE Collections*. Decolonise the Curriculum - the library's role, London, UK. <https://decolonisethelibrary.wordpress.com/decolonising-lse-collections-kevin-wilsonlondon-school-of-economics/>