AN INDICATOR FOR STATISTICAL LITERACY BASED ON NATIONAL NEWSPAPER ARCHIVES[1]

Thilo Klein, Anaïs Galdin and El Iza Mohamedou
Partnership in Statistics for Development in the 21st Century (PARIS21), France

Abstract

This paper develops and reports on a composite indicator for statistical literacy as part of the Busan Action Plan for Statistics (BAPS) logical framework as agreed in the 4th High Level Forum on Aid Effectiveness. The Plan's first objective is to "fully integrate statistics into decision-making". Statistical literacy is a prerequisite to effectively use statistics to inform decisions for planning, analysis, monitoring, and evaluation, thus increasing transparency and accountability. The statistical literacy indicator measures the use of and critical engagement with statistics in national newspapers. The target population are journalists and newspaper readers. This excludes the illiterate population and those without access to print or online media. Articles are from RSS feeds of national newspapers, primarily based on -- but not limited to -- the global news aggregator Google News.

1 Introduction

"There are three kinds of lies: lies, damned lies, and statistics" first proclaimed the British Prime Minister Benjamin Disraeli before the sentence was popularised by Mark Twain. Indeed, statistics frequently had itself discredited because of misunderstanding or mistrust (see Wallman, 1993). In today's complexities of our information society, an understanding of statistical information and techniques has become essential both for everyday life and effective participation in the workplace, leading to calls for an increased attention to statistics and statistical literacy (Shaughnessy and Pfannkuch, 2004; Shaughnessy, 2007). The current phenomenon is dual: progress in the use of statistics goes hand in hand with an increase in the misuses and statistical fallacies (Hooke, 1983). A large body of literature, built by teachers, education researchers, statisticians and professional organisations[2] thus calls for improving and measuring statistical literacy, with a special focus on the student population. Begg et al. (2004), for example, underlined the societal motive behind the call for a greater emphasis on statistical literacy in school curriculum, being that students can become active and critical citizens. Callingham (2007) stressed the importance for students to adopt a critical stance about data, referred to as applying statistical literacy.

The call for statistical literacy has recently been echoed by the international community. The Synthesis Report of the UN Secretary-General on the Post-2015 Agenda, "The Road to Dignity by 2030", called itself for a transformative agenda where we "base our analysis in credible data and evidence, enhancing data capacity, availability, disaggregation, literacy and sharing". It stressed that "the world must acquire a new 'data literacy' in order to be equipped with the tools, methodologies, capacities, and information

1 This paper was developed in collaboration with a task team set up by the PARIS21 Secretariat. We are thankful to the task team members: Kenneth Bambrick and Anthony Higney (DfID-Chair), Maurice Nsabimana (World Bank), Adele Atkinson (OECD), Freeman Amegashie (Afristat), Yanhong Zhang (UNESCAP), Reija Helenius (IASE), Barnabe Okouda (INS Cameroun), Pedro Campos (ISLP), Lisa Bersales (Philippine Statistics Authority), Innocent Ngalinda and Albina Chuwa (East African Statistics Training Centre), Margarita Guererro (UNESCAP/ SIAP), Scott Keir (Royal Statistical Society).

2 In particular, the International Statistical Institute has launched the *International Statistical Literacy Project*.

necessary to shine a light on the challenges of responding to the new agenda". To inform this debate, the PARIS21 Secretariat established a task team for the purpose of developing and reporting on a global indicator to measure the current state and future progress in global statistical literacy. The paper presents the outcome of this consultative process and presents a novel measure of statistical literacy based on the use of and critical engagement with statistics in national newspapers. The use of text mining techniques bridges current data gaps in this area and allows the assessment of statistical literacy of an adult population on a day-to-day basis in more than one hundred developing and developed countries.

The paper is structured as follows. Section 2 reviews the literature. Section 3 presents the text mining methodologies developed to measure statistical literacy and provides a brief overview of the keywords analytic. Section 4 describes the data and presents the results. Section 5 presents robustness checks. Section 6 concludes.

2 Literature Review

Conceptualisation of the notion of statistical literacy

The present paper contributes to a body of literature that addresses the need to give a concrete measure of statistical literacy. Despite an international consensus on the value of understanding data and improving global statistical literacy, there is no general agreement on its conceptualisation. While the need for a common definition of statistical literacy has been recognised (see Ben-Zvi and Garfield, 2004) in the literature, Batanero (2002, p.37) summarises that "we have not reached a general consensus about what are the basic building blocks that constitute statistical literacy or about how we can help citizens construct and acquire the abilities". These definitional issues led to the building of an expanding conception of statistical literacy, from purely conceptual to a more applied concept.

Early work tries to provide a comprehensive definition of statistical literacy. Wallman (1993 p.1), for example, defines statistical literacy as "the ability to understand and critically evaluate statistical results that permeate our daily lives—coupled with the ability to appreciate the contributions that statistical thinking can make in public and private, professional and personal decisions" (see also Trewin (2005) for such kinds of broad, generic definitions). The concept directly introduces both a personal and societal need to develop statistical literacy skills. Callingham (2007) endorsed this definition, underlining it also requires an appreciation of the social context. These studies, however, suffer from the lack of methodological tools that would help to quantify levels of statistical literacy or to identity useful skills and competencies to develop. In response, Ben-Zvi and Garfield (2004) include more active components to the definition of statistical literacy. They define statistical literacy as a set of skills that students may actively use in understanding statistical information -- among them, organising data, constructing tables or being able to work with different varieties of data representations. Most of these definitions are strongly linked to the field of education, identifying statistical literacy as a primary goal and a need for statistics instruction, because "most students are more likely to be consumers of data than researchers" (Garfield and Gal, 1999, p.4). In that regards, Gal (2004, p.1) sees statistical literacy as one of the prominent pre-requisite for participation in society, the "key ability expected of citizens in information-laden societies". His statistical literacy concept involves both cognitive and dispositional components, where some components are common with numeracy and literacy whereas others are unique to statistical literacy. This definition encompasses both critical evaluation of statistics and the ability to express one's opinions or data-related arguments about it. Likewise, Schield (2004) and Watson

(2006) see the ability to question claims in social contexts as a fundamental element to statistical literacy.

Within the nebula of definitions, the concepts of data literacy and statistical literacy are often used without distinction. The present paper adopts the division used by the Oxford Dictionary of Statistical Terms assessing that data literacy can be seen as a subcomponent of statistical literacy, which (Dodge, 2003) defines as “the ability to critically evaluate statistical material and to appreciate the relevance of statistically-based approaches to all aspects of life in general”[3]. In particular, the Australian Bureau of Statistics’ Education Services considers four criteria essential for statistical literacy: data awareness; the ability to understand statistical concept; the ability to analyse, interpret and evaluate statistical information; the ability to communicate statistical information and understandings. This emphasis in the ability to understand and communicate about statistics is recurring in the recent literature: “Statistics requires the basic understanding of statistical concepts…whereas literacy requires the ability to express that understanding in words, not in mathematical formulas” (Watson and Kelly, 2003). Milo Schield (2004, p.9) for instance supports that statistical literacy is “typically more about words than number, more about evidence than about formulas”.[4]

The complexity of the statistical literacy construct, by emphasising the place of critical thinking, contextual understanding and students’ dispositions, offers a real challenge for assessment. Despite these challenges in terminology, several frameworks have attempted to model the features of statistical literacy, focusing mainly on a student population. Our indicator builds on these models to provide a reliable and more widely applicable measure of statistical literacy.

Empirical frameworks

Gal (2004) developed one of the first models evaluating the understanding of statistics by adults. In this model, cognitive and dispositional components interact together. In particular, statistical literacy presupposes the use of five interrelated cognitive elements: mathematical knowledge, statistical knowledge, literacy skills, knowledge of the context and critical questions. An important claim in Gal’s model is that all components leading together to adopt a statistical literacy behaviour constitute a dynamic set of knowledge and dispositions, strongly context-dependent and interrelated entities. The model particularly examines how a person’s dispositions or attitudes toward data and statistics interact with these knowledge bases to motivate a critical thinking about statistics. Once a certain level of statistical literacy is reached, individuals would be able to automatically transfer their skills to evaluating everyday life statistical information they encounter. Gal’s model draws the important implications that anyone lacking these skills is functionally illiterate as a responsible, informed and productive citizen and worker. As suggested by Batanero (2002), the strength of Gal’s model is that it offers a full definition along with the necessary components to achieve statistical literacy, which is particularly for helping policy makers to take decisions at a macro-level of analysis. However, analysing statistical concepts related to this notion requires to use more specific micro-level models (and somewhat a less exigent definition).

A second model, the Statistical Literacy Construct from Watson and Callingham

[3] See also the Australian Bureau of Statistics’ Education Services, who considers four criteria essential for statistical literacy: data awareness; the ability to understand statistical concept; the ability to analyse, interpret and evaluate statistical information; the ability to communicate statistical information and understandings.

[4] Approach equally supported by Biggeri and Zuliani (1999) and Watson and Kelly (2003): “Statistics requires the basic understanding of statistical concepts…whereas literacy requires the ability to express that understanding in words, not in mathematical formulas.”

(2003) builds on the Structure of Observed Learning Outcomes (SOLO) taxonomy developed by Biggs and Collins (1982) to hierarchize statistical thinking into six stages of skills, that can be viewed as a progression of levels of statistical understanding[5]. The strength of this model is that its statistical literacy scale has been widely validated by researchers, based on responses from a large number of students in Australia. At the top two levels of the Watson and Callingham, (2003) construct, students display skills matching the critical-thinking skills of the third tier of the Statistical Literacy Hierarchy in Biggs and Collins (1982). This model of measuring statistical literacy was born to solve the lack of research proposing methods to measure students' progress -- despite statistical literacy being part of the school curriculum.

The differences in approaches between the Gal (2004) and Watson and Callingham (2003) model can partly be explained by the fact that Gal's construct is developed for an adult population while the construct of Watson and Callingham was developed for students. These two main frameworks for statistical literacy are by no means the only ones. Wild and Pfannkuch (1999) proposed a model for statistical thinking in empirical enquiry, built upon statistics education literature and interviews with statisticians and undergraduate students. Reading (2002) relies on the SOLO taxonomy across five areas of statistics[6] to build a "profile for statistical understanding". This methodology, as well as the one developed by Jones et al. (2000)[7], is very similar to the hierarchical model of Watson and Callingham (2003).

As a framework, we adapt ourselves the taxonomy developed in Watson and Callingham (2003). The main methodological contribution of our paper is that it develops text mining methods to measure literacy and critical thinking based on articles from RSS feeds of national newspapers. By relying on text mining techniques, our methodology targets mainly journalists and newspaper readers. This excludes the illiterate population and those without access to print or online media. Moreover, this paper contributes to the existing literature by targeting a population that is not limited to students. The original statistical literacy construct of Watson and Callingham (2003) involves six stages: Idiosyncratic, informal, inconsistent, consistent non critical, critical, and critical mathematical. This taxonomy is developed for students in grades 5 to 10. Our interest is in an adult population and therefore attention is on the top three levels of the taxonomy, as characterised in Table 1.

Level	*Brief characterization of levels*
1: Consistent Non-critical	Appropriate but non-critical engagement with context, multiple aspects of terminology usage.
2: Critical	Critical, questioning engagement in contexts that do not involve proportional reasoning, but which do involve appropriate use of terminology.
3: Critical Mathematical	Critical, questioning engagement with context, using proportional reasoning particularly in chance contexts, showing appreciation of the need for uncertainty in making predictions, and interpreting subtle aspects of language.

Table 1. Statistical literacy construct, adapted from Watson and Callingham (2003)

[5] Boundaries between levels are not rigid.
[6] Namely data collection, data tabulation and representation, data reduction, probability and interpretation and inference.
[7] Proposing four levels of thinking across four key constructs for young children's thinking.

3 Methodology

Data sources

To measure statistical literacy empirically, we turn to references to statistics and statistical fallacies in national newspaper articles that are accessible online, in line with the work done by Watson and Callingham (2003) in terms of scaling. This is essentially for three reasons.

First, and foremost, while there is some gap between journalists' perception of statistics, which is reflected by statistics reported in news articles, and the demand for statistics in the audience, the writing of journalists can be seen as an image for a nation's demand for statistical facts as well as the depth of critical analysis. In any case, in most parts of the world, it largely reflects the nation's consumption of statistical facts as well as the level of critical analysis of statistics offered to a country's population.

Second, newspaper articles are generally available, most of them online, which makes them representative for a country's literate population and easily accessible for text analysis.

Lastly, alternative data sources are either not representative (e.g. Google Trends searches related to statistics; downloads of statistical software packages) or are reported infrequently and/or not comparable across countries (e.g. job categories related to statistics; regional numeracy assessments).

The indicator used is a three-dimensional composite indicator of the equally weighted percentages of national newspaper articles that contain references to statistics at statistical literacy level 1, 2 or 3, respectively, following the scale defined in Table 1. The three levels are not mutually exclusive. For each of the three levels, we obtain the share of documents that match the classification, country per country. An overall measure for statistical literacy is then obtained as the sum over the three shares. Specifically, the methodology classifies keywords used in each article into literacy levels 1 to 3 based on three corresponding keyword lists, so that for each of the 3 levels, there is a different denominator of newspaper articles that is analyzed (see below for a precise description of the keyword analysis). Each keyword list contains different terms referring to statistics and statistical fallacies, and the use of one precise category of keywords by one newspaper article allows for defining one level of statistical literacy.

The empirical instrument is still under construction and the preliminary results described here are helpful to improve the quality of measurement. To establish the validity of the measure, the classification of articles will be further validated by analysts at National Statistical Offices (NSOs).

Text mining techniques

This subsection summarises the keywords used in the analysis, and the sources used to define the appropriate keywords. It also provides examples of keywords defined for each level of statistical literacy. Keywords are derived from major statistical data sources and refer to wide categories of indicator, based on standard internationally adopted by NSOs, International Organisations, books, articles and glossaries specialized in statistics and statistical fallacies (examples are the OECD Glossary of Statistical Fallacies or the Glossary of Statistical Terms by the University of California, Berkeley, for English keywords; or the Glossário Inglês-Português de Estatística for Portuguese keywords). The detailed list of keywords used in the analysis, data source and preliminary results are available at http://paris21.org/literacy.

The study further used the World Bank's WDI database (World Development Indicator) to extend the initial keywords list and added a blacklist of keywords to disentangle ambiguous meaning of acronyms. The reliability and validity of the keyword lists will be further tested during the implementation of validity checks (see below).

Note: Keywords have been translated in all four languages used for the indicator. Text mining techniques, as word stemming, were applied to all keyword lists and news articles before proceeding with the analysis. For articles, stop words were removed and characters are converted to lower case.

LEVEL 1: CONSISTENT, NON-CRITICAL USE OF STATISTICS

Data source: Daily, top 100 news articles from Google News for publishers who
- have registered their RSS feeds with this service,
- publish in either English, French, Spanish or Portuguese

Keywords: articles are considered a good fit for this category if they contain words from one of the following lists:

1. Keywords indicating data sources
   a. word sequences of length two, derived from list of all NSO names worldwide
   b. main statistical data sources, such as 'population census', 'household survey', 'geospatial data', etc (cf. Espey et al., 2015)
2. Keywords indicating a statistical indicator:
   a. GDP, CPI, etc. based on the World Development Indicator database's 'Economy and Growth' category. This list is currently being extended using additional keywords from other categories.
3. Keyword list from statistical capacity building projects

Example: Level 1. consistent, non-critical
- Sentence: "The report indicates tobacco use has increased since the Kenya *Demographic Health Survey* conducted in 2008-09, which found 19 per cent of men and 1.8 per cent of women use tobacco."
- Source: The Star, Kenya

LEVELS 2 AND 3: CRITICAL ENGAGEMENT WITH STATISTICS

Data source: Daily, top 100 news articles from a Google News search for either: 'statistics', 'data', 'study', 'research', 'report'. For publishers who
- have registered their RSS feeds with this service,
- publish in either English, French, Spanish or Portuguese, and

Keywords: articles are considered a good fit for this category if they contain words from one of the following lists:

1. Critical mathematical engagement:
   a. **List of statistical fallacies**: based on books, articles and websites that discuss statistical biases and fallacies
2. Critical non-mathematical engagement:
   a. **List of adjectives to assess the quality of research studies**: based on synonyms and antonyms for 'accuracy', 'reliability' and 'validity' (cf. Pierce, 2008)

Examples:

Level 2. critical

- **Sentence:** "Dr Barres admits a definitive *scientific* conclusion for how these epigenetic changes affect the gene is not yet *scientifically* known."
- **Source:** Citizen Digital, Kenya

Level 3. critical mathematical

- **Sentences:** "Without going to the details of the *statistics*, the final results found […]. *Sample sizes* were calculated at regional level in order to estimate global acute malnutrition with a desired precision of between 2-4 percent with a design effect of 1.5."
- **Source:** Daily News, Tanzania

Limitations

The data source has several limitations that are usefully addressed. First, and foremost, our hierarchy of statistical thinking into three stages of skills (progression of non-rigid levels of statistical understanding based on the SOLO taxonomy) creates a scale that has widely been validated empirically as a measure of statistical literacy. Nevertheless, the indicator is measuring a count of terms specifically referring to each level of literacy, whereas literacy would also need to be tested against the "appropriateness" of the terms used, in context. Therefore, the measure is conditional on the assumption that statistical terms are appropriate for the context they are used in. This assumption is essential to a fully automated process allowing a daily collection and analysis of newspapers articles.

Second, the current implementation is limited to the four most widely spoken languages globally (English, French, Spanish and Portuguese) and thereby ignores local languages. Extending the analysis would require software that allows word stemming and stop word removal in these local languages. An initial analysis of newspapers coverage nevertheless reveals that a vast majority of countries have national newspapers available through their RSS feeds and written in one or several of these four languages.

Third, newspapers and blogs are only a subset of national media. Radio and TV, however, cannot easily be captured in machine readable format. New promising tools, as the Radio Analysis tools[8] developed by Pulse Lab Kampala and the United Nations in Uganda, could maybe fill this gap in the coming years. Radio data could for instance be useful in the future to do a robustness check to see how the use of statistics differs in urban areas – that have access to (online) newspapers – from that in rural areas and illiterate populations. Moreover, automated text analysis does not cover visualized data, such as graphics and tables, an important way of presenting statistics in news media.

Finally, while based on high-level glossaries and internationally acknowledged statistical data sources, the keyword lists used for the analysis are subjective.

4 Results

Scope of the indicator

The purpose of the indicator is to set and monitor targets and report on them annually. Target countries comprise all International Development Association (IDA) borrower

[8] The United Nations Initiative Pulse Lab Kampala is developing a tool to analyse radio content, currently tested in Uganda. For more information on this project, see http://radio.unglobalpulse.net/uganda/

countries, Least Developed Countries, Low and Lower-Middle Income Countries, and the whole of the African continent. The analysis for these 104 countries (as of June 2015) can be usefully extended to Upper-Middle and High Income Countries. This indicator currently covers news articles for about half of the 77 International Development Association (IDA) borrowing countries written in four languages (English, French, Portuguese, and Spanish). Four OECD countries -- France, Mexico, Portugal, U.K. -- are included for reference.

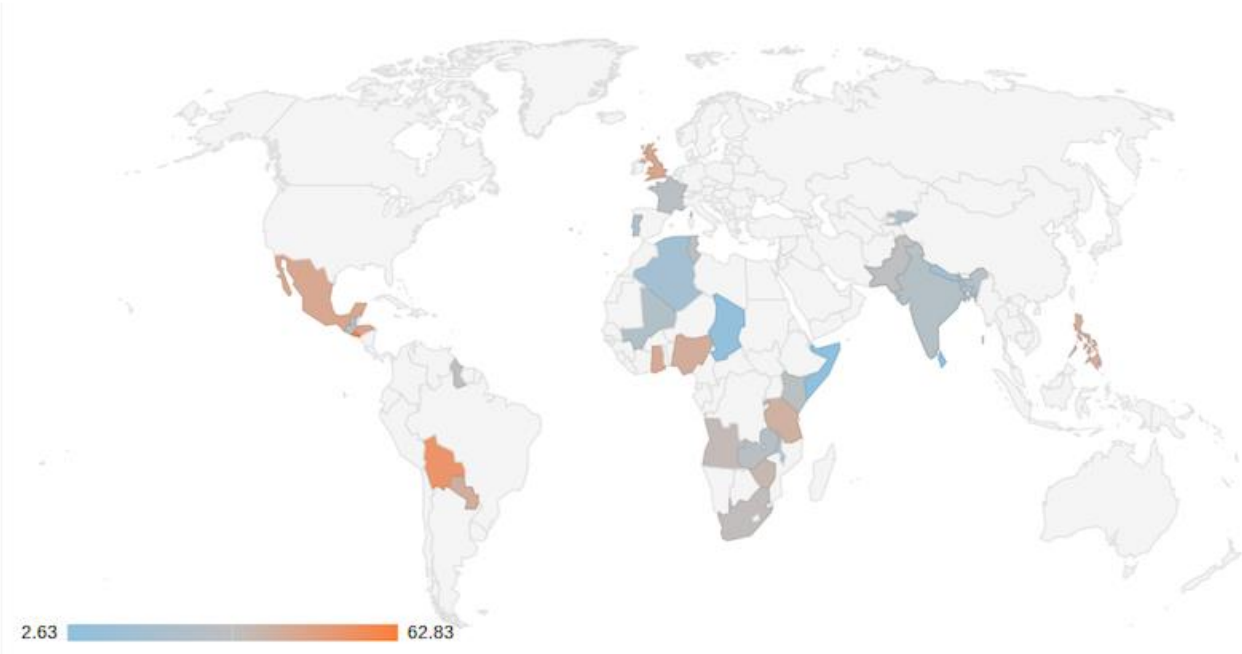

Figure 1. Geographical coverage and scores

Distribution of scores

Starting from 15 April 2016, a total of 8880 articles were analysed for the use of statistics in general news (Level 1). This corresponds to an average of 261 articles per country for the period until 16 June 2016. For Level 2 and 3, a total of 3067 articles with explicit reference to 'statistics', 'data', 'study', 'research' or 'report' were analysed, starting from 15 April. The aggregation score for each country is simply the sum over the three dimensions (ranging from 0 to 300): three-dimensional composite indicator of the equally weighted percentages of national newspaper articles that contain references to statistics at statistical literacy level 1, 2 or 3, respectively, following the scale defined in Table 1. For each of the three levels of statistical literacy, the resulting score gives the percentage of articles that contain at least one search term from the keyword lists defined previously. The score for each level thus ranges between 0 and 100 and the maximum total score over all three levels is 300. The results in Figure 2 are presented by language groups to allow for a direct comparison between countries for which the same keyword list was applied.

The 452 general news articles (corresponding to 5.09 percent of all articles) that cite statistics (Level 1) and the 1094 research-related articles (equivalent to 12.32 percent of all articles) that demonstrate a critical engagement with statistics (Level 2 and 3) can be searched interactively at http://paris21.org/literacy.

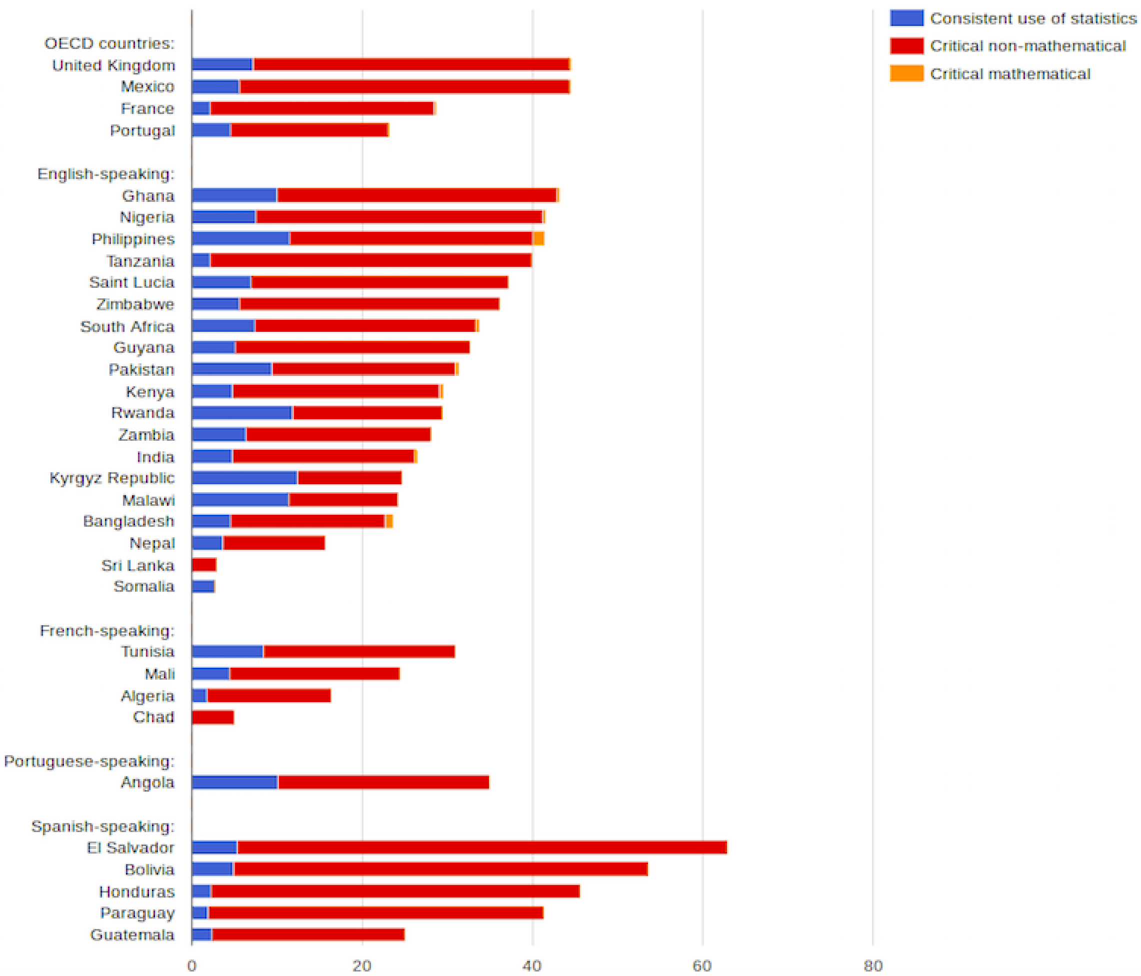

Figure 2. Distribution of scores by country and language group

The results for OECD countries show Mexico joint with the UK on the top of the list and the Philippines come third for the Anglophone developing countries. We offer several explanations for these somewhat unexpected results. First, these two statistical institutes in Mexico (INEGI) and the Philippines (PSA) are very engaged in monitoring the use of statistics by journalists. The PSA, for example, tracks references to their institutions via Google news subscriptions and engages with the media. INEGI reports on the impact and value of statistics based on daily monitoring of newspapers and media resources. A second explanation is the different audience of the main newspapers by country. In the UK, we see a good degree of "yellow press" coverage, which may explain a lower score. Finally, we observe in many of the developing countries that newspapers use the verbatim press releases from statistical agencies without making them digestible for a general audience by removing technical jargon. This points to a weakness in our method, in that it rewards top level keywords related to the critical mathematical category but good journalism should actually avoid technical jargon.

5 Robustness Checks

In this section we carry out robustness checks to test the external validity of our indicator. We compare our results to another available proxy, the **Trends in International Mathematics and Science Study** (TIMSS). TIMSS is a series of international assessments of student achievement, evaluating at least 5,000 students per participating educational system. TIMSS reports every four years on the achievement of fourth and eighth grade students. The first study was conducted in 1995. In 2015[9], more than 60 countries from all categories of economies participated in TIMSS. For most of these countries, TIMSS therefore offers a wide panel dataset.

The implementation of our robustness checks relies on TIMSS in eighth-grade mathematics, with a focus on the sub-section called *Data and Chance*. At grade 8, TIMSS assesses student knowledge in four Mathematics Content Domains: *Number*, *Algebra*, *Geometry*, and *Data and Chance*. The latter represents 20% of the Content Domains for the 2011 and 2015 editions of the TIMSS. The *Data and Chance* section evaluates the abilities of eighth-grade students to understand and extract information from traditional graphic forms, but also from increasingly complicated new forms of visual representation. This section therefore evaluates both the ability to deal with visual display and with the underlying statistics. In particular, the test is designed to test the ability of students to understand misleading or misrepresenting data and statistics. Consequently, this measure aims to capture some very similar concepts to those assessed by our Statistical Literacy Indicator.

Education system	Number	Algebra	Geometry	Data and chance
Korea, Rep. of	618 ▲	617 ▲	612 ▲	616 ▲
Singapore[1]	611 ▲	614 ▲	609 ▲	607 ▲
Chinese Taipei-CHN	598 ▲	628 ▲	625 ▲	584 ▲
Hong Kong-CHN	588 ▲	583 ▲	597 ▲	581 ▲
Japan	557 ▲	570 ▲	586 ▲	579 ▲
Russian Federation[1]	534 ▲	556 ▲	533 ▲	511 ▼
Finland	527 ▲	492 ▼	502 ▲	542 ▲
Israel[2]	518	521	496 ▲	515 ▼
United States[1]	**514**	**512**	**485**	**527**
Australia	513	489 ▼	499 ▲	534
England-GBR[3]	512	489 ▼	498 ▲	543 ▲
Slovenia	511	493 ▼	504 ▲	518 ▼
Hungary	510	496 ▼	501 ▲	517
Sweden	504 ▼	459 ▼	456 ▼	504 ▼
Lithuania[4]	501 ▼	492 ▼	500 ▲	515 ▼
Italy	496 ▼	491 ▼	512 ▲	499 ▼
Norway	492 ▼	432 ▼	461 ▼	513 ▼
New Zealand	492 ▼	472 ▼	483	513
Kazakhstan	479 ▼	506	491	444 ▼
Armenia	474 ▼	496 ▼	450 ▼	376 ▼
Ukraine	472 ▼	487 ▼	476	471 ▼
United Arab Emirates	459 ▼	468 ▼	431 ▼	440 ▼
Lebanon	451 ▼	471 ▼	447 ▼	393 ▼
Malaysia	451 ▼	430 ▼	432 ▼	429 ▼
Romania	448 ▼	477 ▼	453 ▼	429 ▼
Georgia[4,5]	435 ▼	450 ▼	406 ▼	392 ▼
Turkey	435 ▼	455 ▼	454 ▼	467 ▼
Tunisia	431 ▼	419 ▼	426 ▼	398 ▼
Thailand	425 ▼	425 ▼	415 ▼	431 ▼
Macedonia, Rep. of[6]	418 ▼	448 ▼	419 ▼	389 ▼
Chile	413 ▼	403 ▼	419 ▼	426 ▼
Qatar[6]	408 ▼	425 ▼	387 ▼	390 ▼
Iran, Islamic Rep. of[6]	402 ▼	422 ▼	437 ▼	393 ▼
Palestinian Nat'l Auth.[6]	400 ▼	419 ▼	416 ▼	368 ▼
Bahrain[6]	397 ▼	424 ▼	398 ▼	407 ▼
Saudi Arabia[6]	393 ▼	399 ▼	364 ▼	387 ▼
Jordan[6]	390 ▼	432 ▼	407 ▼	379 ▼
Morocco[7]	379 ▼	357 ▼	390 ▼	332 ▼
Indonesia[6]	375 ▼	392 ▼	377 ▼	376 ▼
Syrian Arab Republic[6]	373 ▼	391 ▼	386 ▼	343 ▼
Oman[6]	351 ▼	383 ▼	377 ▼	342 ▼
Ghana[7]	321 ▼	358 ▼	315 ▼	296 ▼
Benchmarking education systems				
Massachusetts-USA[1,4]	567 ▲	559 ▲	548 ▲	584 ▲
Minnesota-USA[4]	556 ▲	543 ▲	515 ▲	571 ▲
North Carolina-USA[2,4]	547 ▲	537 ▲	515 ▲	548 ▲
Quebec-CAN	543 ▲	516	529 ▲	549 ▲
Indiana-USA[1,4]	528 ▲	520	498 ▲	545 ▲
Connecticut-USA[1,4]	527 ▲	510	490	546 ▲
Alberta-CAN[1]	523 ▲	485 ▼	485	529
Colorado-USA[4]	521	512	505 ▲	540 ▲
Ontario-CAN[1]	519	497 ▼	512 ▲	531
Florida-USA[1,4]	517	513	499	528
California-USA[1,4]	492 ▼	509	454 ▼	495 ▼
Dubai-UAE	479 ▼	489 ▼	453 ▼	468 ▼
Alabama-USA[4]	463 ▼	471 ▼	443 ▼	480 ▼
Abu Dhabi-UAE	452 ▼	459 ▼	424 ▼	434 ▼

Table 2. Average mathematics content domain score of 8th grade students, by education system (2011)

We use TIMSS data for the years 2011, 2007 and 2003, for 40 countries and 9 US States for which the results of the *Data and Chance* section are available. Direct comparison of TIMSS data from past years with measures from our Statistical Literacy Indicator allows us to check whether the distribution of TIMSS score matches the one of our scores for the target countries. The data further allows for a matching analysis between more developed regions, via the availability of disaggregated TIMSS results by US States.

[9] TIMSS 2015 scores for the Data and Chance section were not yet officially released by the time of our analysis. Future versions of this paper will incorporate robustness checks using this new set of data.

It could moreover aim to control whether these students, which are now adults, are maintaining the level of statistical proficiency achieved during their studies. This analysis thus compares students and adults' statistical literacy within a country. It may also be seen as a test of the assumption that students who learn to process data manage to transfer these skills in order to interpret and critically evaluate statistical information.

Figure 3 shows a positive correlation between our literacy score and the TIMSS score for a sample of 21 countries. The correlation is significant at the 10%-level as illustrated by the 90% confidence interval of the OLS regression of the TIMSS *Data and Chance* section for the year 2011 on our Statistical Literacy score.

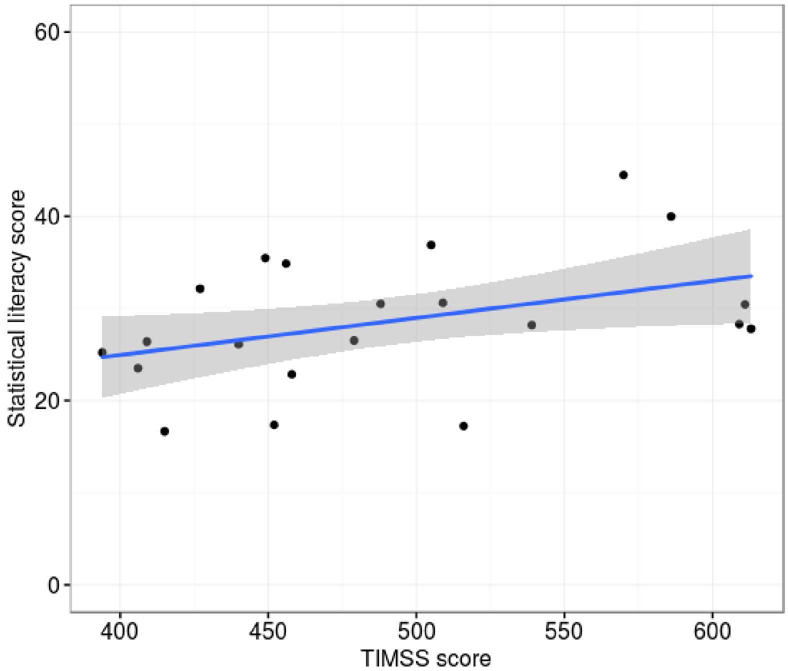

Figure 3. Correlation between Statistical literacy score and TIMSS Data and Chance for the year 2011.

To further establish the validity of the measure, a qualitative evaluation of our methodology is currently being undertaken. The relevance of all newspapers materials identified by our algorithm is to be analysed by partners. Going forward, co-ordination with National Statistical Offices will aim to implement a national monitoring of the indicator on a regular basis.

6 Conclusion

The results presented in this paper are preliminary results. The following steps are currently undertaken to improve the robustness and external validity of the indicator.

- Additional validity checks will be implemented to test representativeness of RSS feeds articles coverage by newspapers by country using figures on average circulations (copies per day). Circulation figures are applied to weigh the contribution of newspapers to the country-level score.

- <u>Two final steps will allow to complete the implementation of the indicator:</u>

1) discussing with NSOs the usefulness of the tool for national monitoring of the use of statistics and dialogue with journalists. NSO's can also give us an assessment of the role of companies and families in terms of statistical literacy.
2) making improvements to the website so that it is publicly accessible, it automates the publication of news in real time and allows visitors to the site to rate articles as good/poor fit for this category.

REFERENCES

Angrist, N., H.A. Patrinos and M. Schlotter (2013). An Expansion of a Global Data Set on Educational Quality: A Focus on Achievement in Developing Countries, *Policy Research Working Paper Series 6536, The World Bank.*

Batanero, C. (2002). Discussion: The role of models in understanding and improving statistical literacy. *International Statistical Review, 70,* 37–40.

Ben-Zvi, D., & Garfield, J. B. (2004). Statistical literacy, reasoning and thinking: Goals, definitions, and challenges. In D. Ben-Zvi & J. B Garfield (Eds.), *The challenge of developing statistical literacy, reasoning, and thinking* (pp. 3–16). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Biggs, J., & Collis, K. (1982). *Evaluating the quality of learning: The SOLO taxonomy.* New York, NY: Academic Press.

Callingham R. (2007). *Assessing statistical literacy: A question of interpretation.* Retrieved from www.stat.auckland.ac.nz/~iase/publications/17/6D1_CALL.pdf

Dodge, Y. (2003). *Oxford Dictionary of Statistical Terms*. Oxford University Press.

Espey, J., Swanson, E., Badiee, S., Christensen, Z., Fischer, A., Levy, M., Yetman, G., de Sherbinin, A., Chen, R., Qiu, Y., Greenwell, G., Klein, T., Jutting, J., Jerven, M., Cameron, G., Aguilar Rivera, A.M., Arias, V.C., Lantei Mills, S. and Motivans, A. 2015. Data for Development: A Needs Assessment for SDG Monitoring and Statistical Capacity Development.

Gal, I. (2004). Statistical literacy: Meanings, components, responsibilities. In D. Ben-Zvi & J. B. Garfield (Eds.), *The challenge of developing statistical literacy, reasoning and thinking* (pp. 47–78). Dordrecht, The Netherlands: Kluwer Academic Publishers.

Garfield, J. (1999), 'Thinking about Statistical Reasoning, Thinking, and Literacy', Paper presented at First Annual Roundtable on Statistical Thinking, Reasoning, and Literacy

PARIS21 (2015). Partner Report on Support to Statistics 2015. Accessed at *http://www.paris21.org/PRESS2015*

Pierce, R. (2008). *Research Methods in Politics*. SAGE Publications.

Rumsey, D. J. (2002). Discussion: Statistical literacy: Implications for teaching, research and practice. *International Statistical Review, 70,* 32–36.

Schield, M. (2004). Statistical literacy and liberal education at Augsburg College. *Peer Review, 6*(4), 16–18.

Shaughnessy, J. M., & Pfannkuch, M. (2004). How faithful is old faithful? Statistical thinking: A story of variation and prediction. *Mathematics Teacher, 95*(4), 252–259.

Shaughnessy, J. M. (2007). Research on statistics learning and reasoning. In F. K.

Wallman, K. K. (1993). Enhancing statistical literacy: Enriching our society. *Journal of the American Statistical Association, 88*(421), 1.

Watson, J.M., & Callingham, R.A. (2003). Statistical literacy: A complex hierarchical construct. *Statistics Education Research Journal, 2(2),* 3-46