Original Article

Data literacy among journalists: A skills-assessment based approach

Ragne Kõuts-Klemm
ORCID: 0000-0002-6422-5546
UNIVERSITY OF TARTU, ESTONIA

DOI: 10.19195/1899-5101.12.3(24).2

ABSTRACT: Datafication brings with it the challenges for journalists to fulfill their historical role as mediators of social processes to their audiences. Journalism has been a rather humanistic field, where journalists tell stories, but do not deal with the analysis and interpretation of numbers. For the current study a methodological tool was developed to measure data literacy among journalists in Estonia. The study confirms that data literacy is acknowledged by journalists as a requirement of future journalism, but their actual skills are still low. Journalists feel more comfortable with data presented in familiar forms. There is a strong tendency that data literacy develops when the skills needed for data processing are in actual use.

KEYWORDS: data literacy, journalism, datafication, statistics, research data.

INTRODUCTION

Datafication changes social and power relations and decision-making processes in societies, as e.g. Schäfer and van Es (2017) indicate in their collection of *The Datafied Society*. The concepts of "data journalism," "computer-assisted journalism," "robot journalism" and others have appeared to point to the related changes in journalism. Different scholars have shown that in the newsrooms we have no common understanding what data journalism should be (Appelgren & Nygren, 2014; Karlsen & Stavelin, 2015; Stalph, 2017). Among scholars themselves data journalism is defined around the core of numerical information — as the usage of numbers and statistics in journalistic reporting (Figl, 2017; Rogers, 2017). Coddington (2015) summarizes the changes under the concept of "quantitative turn" in journalism. The work with statistics, with the results of research, and with data deriving from automated data collection systems will be more and more important to explore journalistically the processes in contemporary societies.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

Many scholars and practitioners see that datafication can save journalism from the crisis, as "the data revolution allows journalism to become transparent and open-source" (Hammond, 2017, p. 412). But on the other hand, Hammond is critical, because the focus of discussion is for him too "post-humanist" — "data journalism is understood as resulting from better data access rather than as being something achieved by the journalist" (Hammond, 2017, p. 408). Without doubt, data has little value if one cannot analyze and interpret it.

In the current study we take a "humanist" approach, since we will focus on the information processing skills among journalists. We select as a case study journalists working in Estonia — in a digital and highly datafied country. The concept of data literacy will be developed and applied for researching data usage skills among journalists. To achieve this aim, the combined methods of individual interviews and a data literacy questionnaire will be adopted for 10 journalists from different media channels and platforms. The current research is explorative in character and will serve as a basis for future comparisons in other datafied contexts.

RECENT STUDIES ABOUT THE USE OF DATA AMONG JOURNALISTS

Data is a broad concept — every piece of information deriving from the environment of human beings that helps to reason and act on them in this environment can be treated as data. In scientific use data means that facts about reality are collected in clearly reasoned and defined ways, with carefully proofed methodology. In the social sciences the differences between quantitative (numerical) and qualitative (textual) data are important. Usually, if we talk in our everyday lives about data, we refer to information presented in numerical form, collected by counting some units or calculated based on algorithms. The current study follows the dominant meaning and restricts the concept of data to statistical and research data. This is in line with many studies on data journalism today (see for instance Coddington, 2015).

There is a growing body of studies that ask how journalists use data. Researchers found in analyzing the content of programs of the BBC some years ago that the use of statistics is limited, mainly mentioning some numbers in passing and rarely giving a context: "Statistics were most commonly used in the coverage of politics, health, business and the economy" (Cushion et al., 2016, p. 5). Researchers concluded that the usage of data is frequently restricted to only very basic levels (BBC Trust, 2016). Another study, conducted by Wihbey (2017) about the use of academic knowledge in journalism reached more optimistic conclusions. Wihbey found that among US journalists, gathered together around the Shorenstein Center, "52 percent reported that they frequently use research studies" and that research is "very helpful in terms of deepening the story context and strengthening story accuracy" (Wihbey, 2017, p. 1279). Studies carried out in the Nordic countries reveal that in work with data the journalists continue their usual work practices (Appelgren & Nygren 2014; Karlsen & Stavelin 2014). Karlsen and Savelin (2014, p. 45)

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

conclude that "the access to data in Norway is perceived as good […] but when the data are analyzed and facts or trends are found, finding a suitable form to present the results can be a challenge."

Journalists mostly deal with "ready-made data" by bodies of experts — by statistical offices, state officials, or by academic researchers. This kind of data is already interpreted, journalists usually have no need to conduct a statistical analysis themselves to find the best interpretation to the patterns the data show them. Critical voices say that journalists depend on the information they get, and often the pieces of information they get are already "framed." Experiments among audiences show that the frames are a powerful basis for audiences to prefer particular interpretations of facts (see Kahneman & Tversky, 1984). Entman (1993) warns that not only audiences, but journalists as well cannot encode the intentions behind the use of particular frames. Recent studies rather confirm the concern that journalists frequently only replicate information provided by data owners, without critically analyzing the interests behind the data (e.g., Simons et al., 2017; Waldman, 2011). A study by Stalph (2017) indicated that journalists use as the main source for data reporting the pre-processed data drawn by domestic governmental bodies. The critical attitude toward data is rather low or even missing. Godler and Reich (2017) confirmed that journalists' belief in the presented information is not even related to the aspect of whether they trust or not the sources who gave the information, but the verification of facts seems to be occasional. "Studies of journalistic verification seem to suggest that the extent of verification in journalism is quite limited" (Godler & Reich, 2017, p. 562). In this habit the journalists are not alone, as their attitudes correspond to the general "high trust in numbers" (Porter, 1995) in modern societies. Godler and Reich (2017) suggest for more "practical skepticism" in the everyday work routines among journalists.

We argue that the results of indicated studies shouldn't be seen only as the low ability of journalists to process and evaluate the quality of data. There are many other intervening factors in the context and work routines of journalism. On the one hand, the current situation can be related to the historical roots of journalism — journalism appeared initially in the form of telling stories and thus belonging rather to the humanistic field. Many schools of journalism started under the umbrella of humanities faculties, and this kind of teaching supports a universal core of professionalism. "The general aims of teaching journalism in most countries are quite similar: students need to know how to express themselves, understand the genres, master the instruments of production and how to handle sources, become familiar with the national professional norms, and so forth" (Hovden et al., 2016, p. 16). Data analyses or research using methods of the social sciences have had less importance in the curricula. Although Weaver and McCombs (1980) have already claimed decades ago that "the use of social science perspectives and methods by journalists is not a recent phenomenon" (p. 477), they must admit that to see the roles of journalists and social scientists more and more in overlapping "may be wishful thinking by some leading journalists who studied the social sciences" (p. 491).

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

Without doubt, the usage of statistical and research data in uncritical ways can derive from specifics of journalistic work processes as well. Journalists have due of the acceleration (Ulferts et al., 2013) today and less time for "back-stage information processing" before publishing (Himma-Kadakas, 2018). The usage of data needs a positive attitude towards statistics and enough time to "play" with data (Karlsen & Stavelin, 2014). Appelgren and Nygren (2014) refer to journalists' opinion that not everybody can be a master in working with data. Thus, the current study challenges the assumption journalists expressed, and asks about their data literacy levels applicable for everyday work.

DATA LITERACY

There are broadly two possibilities to study data literacy — "self-reflective" and "skills assessing." The first method is that used by Wihbey (2017) and by Himma-Kadakas (2018) where a researcher asks the respondents to describe their routines and practices, and sometimes the respondents themselves evaluate their skills and knowledge. Using the first method a researcher relies a lot on the exhaustiveness of the self-reflection processes of respondents. The possibility of insufficient self-assessment is always there.

The second method for literacy analysis is more complicated because it consists of an element of assessment, where the question of the quality of measurement is always present. The instrument of measurement offered by the OECD (2016) to analyze the information processing skills of populations in technologically rich contexts is a good example of the "skills-assessment" model. The OECD (2016) defines information processing as consisting of literacy, numeracy and problem solving in a technologically saturated environment. A study conducted based on this model among European populations — PIAAC (Program for the International Assessment of Adult Competencies) shows "strong associations between proficiency in literacy, numeracy, and problem-solving in a technology-rich environment" (Fridberg et al., 2015, p. 19). For measurement purposes it is needed but in reality it is not easy to draw clear boundaries between performance of different skills.

As the detailed professions-based grouping in the PIAAC database is available only in some countries, the general comparisons between journalists and others are complicated. We can use the data from the Estonian PIAAC-sample[1] to see that journalists employ higher literacy skills than the population in general, but there are no differences in numeracy and in problem solving (see Table 1). The PIAAC study confirms the assumption that literacy, numeracy and information processing skills in tech-savvy environments are higher if they are in actual use (OECD, 2016). The working routines of journalists in Estonia consist of reading and writing, and they have less to do with numbers and calculations.

[1] The group of journalists is in many national samples too small to analyze the differences between journalists and other professions. In Estonia it was small too, but their amount represents their proportion among other professions in the Estonian job market.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

Table 1. Means of indexes among journalists and total population in the Estonian sample

Index	Journalists	Population	Pearson's *R*	Significance
Index of use of reading skills at work (prose and document texts)	4.17	2.82	0.47	0.000
Index of use of writing skills at work	3.50	1.92	0.57	0.000
Index of use of numeracy skills at work (basic and advanced)[2]	2.44	2.40	Not relevant	
Index of use of ICT skills at work	3.78	2.94	Not relevant	
Index of learning at work	3.67	2.82	Not relevant	
Index of readiness to learn	3.61	2.65	0.35	0.003

Note: *N* = 7632, PIAAC study in 2011, calculations by the Author.[3]

Source: Author.

If we talk about data literacy, the component of numeracy is relevant. Numeracy is the core of data literacy, but it would be rather narrow to restrict data literacy only to numeracy. Experiments carried out in psychology show that data usage has a strong factor of reading and interpreting skills as well — when the numerical data are presented in context or are already "framed" (Kahneman & Tversky, 1984) not only the ability for mathematical calculations is needed. The receiver should be able to recognize attempts to "lie with statistics" (Huff, 1954).

In the current study the data literacy for the profession of journalism will be thus defined as the five steps data-processing model: a journalist is able (1) to find the data, (2) to evaluate the data quality, (3) to interpret the data in the context, and (4) to present the data by journalistic means, (5) taking into account the needs and capacities of reception by the audiences.

The aim of the study is to explain data usage among journalists in their work routines and to highlight the gaps between the actual data literacy and the needs of the datafied society. The study sets two research questions:

RQ1: What are the attitudes of journalists toward statistical and research data? How much they do use the data?

RQ2: Do they have the skills to process the data presented in different ways, as mathematical, logical and research data? Do they practice a "practical skepticism" in working with data?

[2] The differences between different respondents are not big; in the Estonian sample standard deviation is one of the lowest — it means that numeracy is almost the same among different groups — a rather different situation from that in other countries like the US, France or Israel (see https://www.oecd-ilibrary.org/education/skills-matter_9789264258051-en, Figure 1.2.).

[3] I am grateful for the PIAAC data to the Estonian Ministry of Science and Education (especially to the study coordinator Mrs. Aune Valk).

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

METHODOLOGY AND SAMPLE

While the aim of the study is twofold — to explain the use of data in the journalistic production process and to assess data literacy among journalists, the study combines two methods. Firstly, individual semi-structured interviews were carried out where the questions about the use of data in everyday practice, their availability, and main sources for getting the data were asked. The interviews had an explanatory role, to give context and information about the attitudes the respondents have toward data. The second part of the study consisted of fulfilling the requirements of the questionnaire by the respondent in "thinking aloud" — a method widely used in psychological studies to model the cognitive processes (van Someren et al., 1994). Registering the reactions to the questions and the verbalized thinking processes enable us to analyze the convenience in dealing with the data.

The main methodological challenge for the study was the data literacy questionnaire. The questionnaire was worked out during pilot testing and had in the beginning another structure and somewhat different questions. The reactions during the pilot study revealed that it is highly important how numerical information is presented to the respondents. The wording has to be clear, concrete and preferably in the form the respondents use in the journalistic production process — numbers given in the context as news. Thus, the final questionnaire was developed to balance the familiar and non-familiar presentation of data for the respondents with the aim of not causing a refusal to participate in the study.

Researchers decided on a final questionnaire with ten questions that represented four types of questions: (1) basic numeracy — "school-math," (2) logical and probability questions without framing; (3) logical and probability questions with framing (based on the example of Kahneman & Tversky, 1984; Entman, 1993); (4) presentation of research data (presented in journalistic as well as in sociological formats) (see the example questions in the appendix). The questions in the questionnaire were given in an ungrouped way, starting with two rather simple questions and after that some more complicated questions. Eight questions had variants of answers where a respondent had to choose the correct one. Two questions were open-ended, comments to the presented information were asked. One of the open-ended questions was a short news story that presented the study results in journalistic form and the other presented a table about the latest changes in the ratings of Estonian TV channels. The aim of adding these questions to the questionnaire was to collect information about "reading" the data — does a respondent make social-scientific or journalistic remarks about the presented information, is the overall mood toward the data critical, interpretative, etc.

As the sample 10 journalists from different media houses in Estonia were selected, the main criterion for selection was that they work as reporters who collect information and produce the news stories themselves (the editors were not included in the sample). The sample consisted of journalists with varied work experience,

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

they represent the main channels on the Estonian media landscape (local and national, commercial and public service media) and they cover different topics (from soft to hard news topics) (see Table 2). The respondents do not identify themselves as data journalists.

Table 2. Sample of journalists interviewed and tested (the information in the table was given by the respondents during the interviews[4])

No	Media organization/ channel	Work experience as a journalist	Topics covered[5]	Education
R1	Local print and online outlet (commercial)	18 years	All possible local topics	HE in journalism
R2	Quality national print and online outlet (commercial)	10 years	Politics, economy	HE in language studies
R3	Popular national print and online outlet (commercial)	9 years	Human interest stories	SE
R4	Quality national print outlet (commercial)	10 years	Investigative journalism, social topics	HE in journalism
R5	Public radio, national	8 years	Politics, economy, finances, public bodies	HE in journalism
R6	Public media, online	4 years	Science, technology	HE in technology studies
R7	National business paper (print and online)	3 years	Economy, finances, entrepreneurship	SE
R8	Online news media, commercial	2 years	Different topics	HE in governance studies
R9	Public media, online	3 years	Daily news from Estonia	HE in language studies
R10	Regional paper, print and online	8 years	Daily news from Estonia	HE in genomics

Note: HE — higher education; SE — secondary education.

Source: Author.

[4] The interviews lasted from 28 to 50 minutes and were conducted without being disturbed by a third party.

[5] Many media outlets in Estonia are rather small and the specialization to one topic among journalists is not usual.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

We asked about respondents' usage of statistical and research data for the news production and it appeared that the importance of data for single journalists was quite different. The heaviest users of different data and databases are respondents R2, R4 and R6. They treat the data processing as unavoidable and as a steadily growing part in the future journalism. Although they use data a lot, they expressed different self-confidence in dealing with data. The respondents R2 and R6 evaluated their skills as rather good and enough for today's needs; the respondent R4 said that more knowledge about the data collection methods and data mining would be needed, the respondent personally feels.

Almost the non-user of data is only one respondent (R3) in the sample who explained that the personified news stories or features usually don't require the use of data. The respondent works currently for a national tabloid, but based on former work experience in local media the respondent remembered that more data skills were not needed either.

Other respondents are moderate users of statistical and research data — they sometimes need to ask additional information to give a broader context to the story. Then they ask for data from statistical offices and from state authorities, rather rarely from public opinion companies or from academic research. They all said that they have particular data bases and web channels they study regularly, related to the topics they dominantly cover — in some ways they are specialized to particular data.

The sample appears to be heterogeneous in the dimension of the data use practices of the respondents. Next, we present their attitudes toward the usage of statistics and research data more precisely.

RESULTS

Attitudes toward usage of statistics and research data

The respondents expressed in the interviews the common view that in today's world data processing skills are necessary, and this viewpoint was not related to the fact how much data a respondent personally uses for his/her daily work. As the study by Wihbey (2017) among US journalists showed, expert knowledge is more frequently drawn from academic research by journalists who work on national level and less by journalists on the local level (2017, p. 1277). In our sample we cannot find confirmation of the distinction.

> Data are needed in journalism. Our editors ask me more and more to have supplementary, factual information. Even if I make a feature where only one person stands in the focus. [Interview, R3]

> I think that even in working for local media data literacy is needed. If a local journalist has the next year's budget of local government on the table, she/he has to be able to understand what the numbers mean in the document. [Interview, R2]

Respondents who use data only modestly or not at all evaluate their data literacy as enough for their daily work needs. Most often they mentioned that they ask for

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

data from communication specialists or public relations officers from particular ministries, state bodies or from statistical offices. If they do not understand the information they get, or when some numbers are confusing, they have continuing communication with the data-providers until they have made clear for themselves what the data mean and how they can interpret it.

> I get data quickly and well. For example the Estonian Rescue Board or the Police and Border Guard Board respond speedily and their numbers are easily understandable. Usually they answer exactly the question I ask. Sometimes the Estonian Health Insurance Fund or the Statistical Office gives me raw tables — these are confusing. Then I have to ask again. [Interview, R1]

> I do not think that the availability of data in Estonia is a problem. On the contrary, the new data protection directive[6] has made the life of journalists complicated. We can easily find statistics and general data, but it is almost impossible to find and get information about a particular person! But a story based only on statistics is not interesting for the reader. A reader wants people! [Interview, R3]

Respondents who can be labeled as the heaviest users of data found that their data literacy is sufficient, only one of them was critical as regards her skills in processing the data.

> It is always good to be better. I know that I have to learn more about data processing. [Interview, R4]

In the case where help is needed, the respondents say that they get support from the data specialist working in the editorial house or from a colleague who is better equipped with data processing skills. It can be concluded from the information the respondents gave that at least the bigger media houses in Estonia have employed a journalist who is acknowledged as "a master of data processing" — who is equipped with the information about the existence of public and private databases, who has contacts with data owners and knows where to find and how to download the data.

Under the data usage skills data visualization skill was not touched on at all in the interviews, by none of the respondents. Probably we can explain it if we know that graphs, tables and other visuals are not made by the reporters themselves, but usually by a person employed just for visualization and layouting.

> I do not know that we had real data journalists in Estonia. A real data journalist is […] hmm […] who does not search for the data that confirms the story one already has, but who analyzes masses of data and writes the story from "hearing" the data, what the data is telling him/ her. [Interview, R2]

The respondent refers to the new data-driven paradigm in data science that scholars dealing with "big data" have defined as a proper way to analyze the data available thanks to the technological developments and "revolutions in measurement" (see Kitchin, 2014). But the interviewed journalists were hesitant when it comes to journalism in Estonia having enough resources to follow this new paradigm.

If we analyze the interviews from the perspective of the function the data have for respondents' journalistic production, then it was common for all the respondents

[6] The respondent refers to the General Data Protection Regulation by the EU. See more: https://www.eugdpr.org/.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

to see data as a supplement for journalistic production. Statistical and research data give a broader context to support the conclusions they already have, similarly in Wihbey (2017). Data verifies and legitimizes the story.

> The editors ask always to have more data, more factual information to make a story more multidimensional. [Interview, R1]

> The data as a pure basis for the story is rather rare in my work. Usually the data I am using gives supplementary deepness to the story I already have. [Interview, R2]

One respondent even expressed an opinion that a story based on statistics is boring, that news stories are full of numbers today, and that readers will complain about too much data.

> Statistics is boring. The story based on statistics is desiccate. Nobody will read it. [Interview, R1]

So, the main discrepancy from the interviews appears to be the inconsistency between the normative view, between attitudes toward data and their actual use — respondents claim that data processing skills are more needed, but they personally are satisfied with their skills to process the data (even if the skills are rather modest or even deficient).

Data processing skills

To analyze data processing skills among the journalists, we had four types of questions in the data literacy questionnaire and we present the analysis of the types separately. We take into account the amount of correct answers in every category and the comments given by "thinking aloud."

Firstly, if we analyze the amount of correct answers, it is surprising that the variance between the respondents in answering was rather low (*Stdv*. = 0.002 on the binary scale, $p < 0.001$). It corresponds to the results of the PIAAC study, where the Estonian population appears to be rather homogeneous in their numeracy skills — the variance in numeracy is similar to the Slovak and Czech populations, and significantly lower than e.g., among Swedish, Danish or German populations (OECD, 2016, Figure 1.2). But the respondents in the current study had differences in answering different types of questions — some of the questions were easier for them, some of the questions were rather more complicated.

The easiest questions for the respondents were the "school-math" questions, where the number of correct answers was 73% (Table 3). With one exception, the respondents answered these questions quickly and without using some helping devices (even no paper and pencil). Nevertheless, the first reactions when the respondents saw the questions were surprising: "Oh, this awful math!" [Interview, R6] or "I am not good at calculations! Shall I use a calculator?" [Interview, R1]. But this form of information is something they say they do not need in their everyday lives — they do not calculate numerical information themselves. But they all have learned it for 12 years in the state school system (they all have at least a secondary

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

level of education). As the interviews showed, the respondents use typically ready-made data for story-telling. One of the respondents even expressed confusion, if the interviewer asked about the trustworthiness of the ready-made data:

> I have never, ever thought about the possibility that the numbers could be wrong in the press releases or in the statistical tables! … Hmm, maybe I should check the numbers? [Interview, R1]

Table 3. Results from the data literacy questionnaire

Type of questions	Number of questions	Amount of correct answers
T1: "School-math"	3 questions	73%
T2: Logical and probability calculations presented without framing	2 questions	60%
T3: Logical and probability calculations presented with framing	2 questions	50%
T4: Presentation of sociological data	3 questions	Cannot be calculated

Source: Author.

Logical and probability calculations presented without framing were the second easiest type of question, but the amount of correct answers was lower than in the dimension of basic numeracy. The most confusing were the probability calculations, where the respondents relied on their "feelings" and not on the mathematical calculations. For example, some of them refused to answer the following question: Which probability is higher — A or B? A. To get in throwing the obverse of the coin up; B. That today is a rainy day if yesterday was rainy.

> It is a matter of coincidence. Nobody knows. It can be sunny at the moment and the next minute you can have rain from a clear sky. [Interview, R2]

Similar experience-based comments were made by other respondents. Thus, if we follow the theory by Kahneman (2011), the "fast system" of the human brain was working if the respondents answered the questions. No respondent ratiocinated in using the terminology of probability calculations — that in reality we have dependent and independent events and their probabilities are different, calculated based on different mathematic formulas.

Even more complicated for the respondents were the logical and probability calculations presented with framing. The number of correct answers to this type of question was only half.

When the information was presented in the form offered by Kahneman & Tversky (1984), where the decision for the correct answer should be made seemingly as a moral decision, the respondents felt themselves to be under pressure. They were asked to choose between solutions where "200 people will be saved" or where "with

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

one-third probability 600 people will saved and two-thirds probability no one will be saved." Mathematically the probabilities are the same.

> It is an ethical question! Of course, I will choose the answer where more people will be saved. [Interview, R3]

To this framing question there was in the questionnaire pre-given the variable to not prefer any solutions, too. Seven respondents from ten selected this variable for answering. But in answering these questions the skepticism of journalists appeared more clearly and it was rather usual that they refused to answer.

> It is suspicious. There are so many emotions in the question. I think that I should check the information. I cannot answer, I do not know yet. [Interview, R6]

The process of answering the fourth type of question — the presentation of research data — indicates that the questions were for respondents seemingly the easiest. They expressed the feelings of being on solid ground in reading the information enriched by data and getting the picture about the contexts. They commented on the information presented in this form in three ways: using (1) professional journalistic arguments (most of them), (2) knowledge about the methodologies of the social sciences (a few respondents), or (3) based on common sense. As an example of the first, they wanted to verify sources, re-formulate the title of the story, or sought the proper channel for the story — "as a piece for yellow press" [Interview, R1]. If the study results were presented in the form of news, the respondents paid attention to the following aspects: who published the study (*Cosmopolitan*), are the results convincing, how big the sample was.

> I don't know. It seems confusing for me. What does it mean — "a woman of average attractiveness"? How can they measure average attractiveness? Do we have a scale for attractiveness? And the sample — 14,000 people — it seems so big. Usually we have much fewer people in the samples, as I know. But firstly, I have to check, do the researchers exist at all in reality? [Interview, R3]

As an example from the methodological comments:

> I cannot understand how make up was measured. How is it possible to correlate the quality of grooming — what is the scale?! — with the time dedicated to or money spent on grooming?! [Interview, R6]

A respondent who read the data table about the ratings of the channels tried to find the causalities behind the changes in the ratings:

> Hmm, I have to remember what the channel had in their program in March that the ratings are so much higher than in other months. [Interview, R1]

But a concerning tendency appeared in this type of question as well — the evaluation of the presented information does not extend the "stock of knowledge" of the respondents. The testing indicated that the respondents knew pretty well the statistical concept of correlation, but in the presentation of data the questions about the methodologies of data collection or calculation of results were rarely questioned.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

The small differences in numbers were interpreted pretty quickly as significant and the causalities were sought.

The main conclusion of the study from this type of questions is: if journalists have more in-depth experience with sociological methods or statistical concepts, they use it in evaluating the quality of information more often than those who do not. But if they have no knowledge — the questioning or criticism toward the presented information will not appear.

CONCLUSIONS

The current study develops the skills-assessment approach to data journalism. The results of the study from the interviews, in combination with completing the data literacy questionnaire, support the conclusions from earlier studies. Statistics or research data helps journalists in deepening their story context, a finding by Wihbey (2017) as well. Second, the critical attitude toward the ready-made data among journalists is rare (Simons et al., 2017; Stalph, 2017; Waldman, 2011) — and this is really concerning. When state authorities, public opinion firms, or statistical offices present some data in their press releases and approaches to the press, journalists do not put the accuracy of the data under question. Even if it is a commercial enterprise which presents their financial data, this data is rarely questioned by most journalists. The same is true if we talk about research data — the criticism toward sociological data is rather weak. Journalists hardly question the methodologies of the studies, interpretation possibilities of the results of statistical analyses, or causal relations between phenomena in reality. One part of the interviewed journalists was eager to find quickly from their personal experiences or from "common knowledge" the explanations to the patterns expressed in data.

But the main conclusion from the study is important for teachers of journalism and for media houses as well. Journalists develop data literacy with a lot of practicing. Even if they are hesitant firstly toward the school-math testing, they show high scores in correct answers — all the respondents "trained" to think in mathematical formulas at least 12 years in Estonia. Probability calculations did not belong for a long time to the standard school program and their practicing has been lower — just as the results of the probability calculations. Literacy in research data is even more diversified — more background knowledge and contacts bring advanced literacy with it. In the sample, journalists dealing with rather "hard topics" like politics, the economy, finances, science, etc. are more data literate, since the fields themselves are pretty datafied. Without better knowledge of "spirit of data" they cannot cover those topics comprehensively. The study shows that data literacy is strongly related to the everyday work practices of journalists (see also Cushion et al., 2016). This corresponds to the PIAAC results (OECD, 2016) as well — a master will be the person who practices more.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

The interviewed journalists themselves said that some of them have more interest in data processing than others. The data literacy needed in newsrooms will be a set of skills developed highly only by a division of journalists. The developments in the journalism job market have different directions open — on the one hand, the job market can be more functionally differentiated in the future, but the opportunity to ask from all journalists to be multi-skilled is an option as well (see Himma-Kadakas, 2018). In a datafied society it is easy to argue in line with the normative expectations that all journalists have to demonstrate high data literacy levels. Research supports this direction of development as well. For example, Appelgren and Nygren (2014) see outcomes from investments into data knowledge and skills in Swedish newsrooms — "an increase in the quality of projects involving data analysis, the creation of more in-depth journalistic projects with limited resources and a strengthening of the role journalists play as gatekeepers" (2014, p. 403). But when the tendency of growing job differentiation in media organizations is true, then it can bring with it a counter-trend, where journalists have a higher specialization regarding the skills and data journalism will be an effort for team-work. The expectations toward multi-skilled journalists are high, but reality brings more nuances into the picture.

In any case, data literacy research needs further development. The method used in the current research was time-consuming and cognitively demanding for the respondents and it enabled only a small sample of journalists. The combination of testing and thinking aloud is valuable to relate actual skills with the qualitative insights into the data use practices by journalists. The current study can be treated as a conceptional case-study, conducted in a datafied culture and thus offering insights into the probable future developments and challenges in the research of journalists' data literacy. Nevertheless, data literacy research needs a cross-country effort to obtain more solid instruments for comparative research.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

APPENDIX

Types of questions in the data literacy questionnaire

Type of questions	Example of a question in the questionnaire	Researcher's comments
Basic numeracy (the "school-math")	Question: Please calculate the percentage of the number 16 from the number 64. Variants: 10%, 25%, 40%, 45%	Questions derived from the math study books for 4th and 8th grade; based on the example of PIAAC, for numeracy see "Sample questions…").[7]
Logical and probability tasks without framing	Question: There are flowers on a lake. An area with flowers will be two times bigger every day. If the flowers extend over the whole lake with 48 days, then over how many days will one half of the lake have flowers? Variants: 24 days; 47 days Question: Which probability is higher: A. To have after the throw the obverse of the coin up; or B. That today is a rainy day if yesterday was rainy? Variants: A; B; probabilities are the same	Questions derived from math quizzes for students in the 10th grade.
Logical and probability tasks with framing	Question: Imagine that a country is preparing for an outbreak of an unusual Asian disease, which is expected to kill 600 people. Two alternative programs to combat the disease have been proposed. Assume that the exact scientific estimates of the consequences of the programs are as follows: If program A is adopted, 200 people will be saved. If program B is adopted, there is a one-third probability that 600 people will be saved and a two-thirds probability that no one will be saved. Which of the two programs would you favor? Variants: I will favor A; I will favor B; it makes no difference because the probabilities are the same	Exact wording derives from Kahneman and Tversky (1984, p. 343), translated by the author.
Presentation of sociological data	A short news story about correlation between make up and income for women (see: https://www.glamourmagazine.co.uk/article/women-who-wear-makeup-get-paid-more).	The news was published in Estonian news media, the translator is unknown.

Source: Author.

[7] PIAAC defines numeracy as the ability to use, apply, interpret, and communicate mathematical information and ideas. It is an essential skill in an age when individuals encounter an increasing amount and wide range of quantitative and mathematical information in their daily lives ("Sample questions..."). The numeracy levels in participating countries can be found in https://www.oecd-ilibrary.org/education/skills-matter/average-and-variability-of-numeracy-scores_9789264258051-graph2-en (accessed: 18.03.2018).

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

REFERENCES

Appelgren, E., & Nygren, G. (2014). Data journalism in Sweden. Introducing new methods and genres of journalism into "old" organisations. *Digital Journalism, 2*(3), 394–405.

BBC Trust. (2016). *BBC trust impartiality review: Making sense of statistics.* UK. http:/downloads.bbc.co.uk/bbctrust/assets/files/pdf/our_work/starts_impartiality/audience_reasearch.pdf.

Coddington, M. (2015). Clarifying journalism's quantitative turn: A typology for evaluating. *Data Journalism, Computational Journalism, and Computer-Assisted Reporting. Digital Journalism, 3*(3), 331–348.

Cushion, S., Lewis, J., Sambrook, R., & Callaghan, R. (2016). *Impartiality review of BBC reporting of statistics: A content analysis.* Cardiff: Cardiff University.

Entman, R. (1993). Framing: Toward clarification of a fractured paradigm. *Journal of Communication, 43*(4), 51–58.

Figl, B. (2017). *Bigger is not always better: What we can learn about data journalism in small newsrooms.* Reuters Institute Fellowship Paper. Oxford: University of Oxford.

Fridberg, T., Rosdahl, A., Halapuu, V., Valk, A., Malin, A., Hämäläinen, R., …, & Mellander, E. (2015). *Adult skills in the Nordic Region: Key information-processing skills among adults in the Nordic region.* Copenhagen: Nordic Council of Ministers, TemaNord.

Godler, Y., & Reich, Z. (2017). Journalistic evidence: Cross-verification as a constituent of mediated knowledge. *Journalism, 18*(5), 558–574.

Hammond, P. (2017). From computer-assisted to data-driven: Journalism and Big Data. *Journalism, 18*(4), 408–424.

Himma-Kadakas, M. (2018). *Skill performance of Estonian online journalists: Assessment model for newsrooms and research*. Dissertationes de mediis et communicationibus universitatis Tartuensis. Tartu: University of Tartu Press.

Hovden, J. F., Nygren, G., & Zilliacus-Tikkanen, H. (Eds.). (2016). *Becoming a journalist: Journalism education in the Nordic countries.* Gothenborg: Nordicom.

Huff, D. (1954). *How to lie with statistics.* New York: W.W. Norton & Company.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kahneman, D., & Tversky, A. (1984). Choice, values, and frames. *American Psychologist, 39*, 341–350.

Karlsen, J., & Stavelin, E. (2014). Computational journalism in Norwegian newsrooms. *Journalism Practice, 8*(1), 34–48.

Kitchin, R. (2014). Big Data, new epistemologies and paradigm shifts. *Big Data & Society, April–June*, 1–12. https://doi.org/10.1177/2053951714528481.

OECD. (2016). *Skills matter: Further results from the survey of adult skills*. OECD Skills Studies. Paris: OECD Publishing.

Porter, T. M. (1995). *Trust in numbers: The pursuit of objectivity in science and public life.* Princeton, NJ: Princeton University Press.

Rogers, S. (2017). Foreword. In J. Mair, R. L. Keeble & M. Lucero (Eds.), *Data journalism: Past, present and future*. Bury St Edmunds: Abramis Publishing.

Schäfer, M. T., & van Es, K. (Eds.). (2017). *The datafied society: Studying culture through data.* Amsterdam: Amsterdam University Press.

Simons, M., Tiffen, R., Hendrie, R., Carson, A., Sullivan H., Muller, D., & McNair, B. (2017). Understanding the civic impact of journalism. *Journalism Studies, 18*(11), 1400–1414.

Stalph, F. (2017). Classifying data journalism: A content analysis of daily data-driven stories. *Journalism Practice*. https://doi.org/10.1080/17512786.2017.1386583.

Ulferts, H., Korunka, C., & Kubicek, B. (2013). Acceleration in working life: An empirical test of a sociological framework. *Time & Society, 22*(2), 161–185.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019

van Someren, M. W., Barnard, Y. F., & Sandberg, J. A. C. (1994). *The think aloud method: A practical guide to modelling cognitive processes*. London: Academic Press.

Waldman, S. (2011). *The information needs of communities: The changing media landscape in a broadband age*. Washington, DC: Federal Communications Commission.

Weaver, D. H., & McCombs, M. E. (1980). Journalism and social science: A new relationship? *Public Opinion Quarterly, 44*, 477–494.

Wihbey, J. (2017). Journalists' use of knowledge in an online world. *Journalism Practice, 11*(10), 1267–1282.

Central European Journal of Communication Volume 12, No 3 (24), Fall 2019