ABSTRACT

‘Fake news’ is now embedded as an everyday term in the media. This has significant consequences for educational curricula. In this paper, we explore the broader context of the misrepresentation of data and the critical thinking and discernment skills necessary to detect it. Increasing ubiquity of data sources and the growing prominence of big data are key developments alongside innovations in digital technology such as analytics for teaching and learning, virtual reality, augmented reality, the Internet of Things, and artificial intelligence. Such innovations in technology are all promising for education; however, these innovations are all taking place within a context of a growing data-driven era. For two decades now, we have been hearing about ‘21st century skills’ and core competencies such as collaboration, critical thinking, communication, and creativity together grouped with digital technology skills citizenship skills. When it comes to ‘digital literacy’ or ‘digital literacies’ there has been a moving feast of options and models. With all the eye-catching infographics and data visualization tools available we are needing to develop better discernment skills – spotting fake news and irresponsible reporting. In other words, we need skills in 'data literacy' and the key question we explore in this short paper is: in what ways can the k-12 curriculum be developed to scaffold data literacy? In contextualizing the problem, we provide a few examples of leading practice and suggest some directions to explore further.

Keywords: Data, Data Literacy, Fake News, Critical Thinking, Curriculum

INTRODUCTION

In 1984, the editors at the Orange County Register changed the colour of Los Angeles sky from in all their outdoor photos of the summer Olympics from smoggy dull to sky blue. In 1998, National Geographic showed in an advertisement polar bears in Antarctica (using a photoshoot from Ohio Zoo). Towards the turn of the century many media outlets were doing similar things, for example, New York Daily (8th September 2000) carried a photoshopped
image of American President Bill Clinton and Cuban president Fidel Castro shaking hands, whereas in reality the hand shake was never photographed. The magazine ‘protected’ itself by and labeling it as ‘photo illustration’. Since then we have come a long way. Data, mobile phones and Internet are ubiquitous and are shaping the individual perception and public policy, as is evident by the US election of 2016 and reports of Russian interference. Such influence is pervasive in all walks of life, and often with no regulatory structures in place.

Data establishes our relationship with reality. Fake news is purposely designed to get this relationship disrupted with the aim of changing realities in the heads of readers. Faking it with data is a subtext of all information that we receive in the era of post-truth. A key question arises: when it is increasingly easy to create lies, fake reports and design deceptions, what needs to change in our educational and real life to counter and upskill us against it? Thus, in pursuing this agenda we have formulated the following research question: In what ways can the k-12 curriculum be developed to scaffold data literacy?

In the traditional curriculum, the scholarship in Mathematics and Science learning areas is mostly based on narrow and closed processes. The processes, that aims to find ideal solutions to (real or imaginary) problems. The aim thereby of this kind of education is to find the answers without questioning assumptions. The mathematics of data within the above context is completely different, as it is open ended, it is first and foremost about identifying questions instead of answers and hence teaching it is complex and multifaceted.

‘Data is too big to be left to the data analysts’ is displayed on the back cover of a recent book discussing anthropology and technology (Bell & Gregg, 2015). It has long been realised that numerical content carried by data is not independent of people or medium. Understanding data requires us to know some context like the who, what, when, where, how, and why aspects (Best, 2004). It is imperative that our future generation understands and be skilled in at least the basic premises of data and such associated questions, and this cannot be done within the parameters in which data skills are being taught currently.

Within our current teaching paradigm, a thorough and deeper understanding of statistics beyond computations and software commands is not yet developed to include ‘data literacy’. Instruction and experimentation with data in the school curriculum is not necessarily taught in ways that facilitate students to see beyond arithmetic and rote answers to explore how biases and false information can or may possibly be imbedded or identified.

Each discipline is now facing the difficult challenge of making selections from content choices due to accelerated pace of ‘knowledge’ arrival. Mostly it is based on big data correlation algorithms – where causation, the pattern finding through scientific theory, or the mathematical formulas are no longer suggested or required (Mayer-Schönberger & Cukier, 2013). This has serious implications on cognitive development as ‘scientific’ advances can be made without looking for causation or actual reasons since ‘correlation supersedes the causation’ (Anderson, 2008). For example, would it mean circumstantial evidence will be the only evidence required to draw conclusions and take decisions in future?

**METHODOLOGY**

This conceptual paper is the result of synthesis of issues emerging from a literature review in which the literature spans theoretical and research literature as well as public policy and
curriculum frameworks, principally from Australia. The focus of the research has been notions of ‘data literacy’ and how this is taught.

**TERMINOLOGY**

*Data literacy* has been defined in various ways (Mandinach & Gummer, 2013). An early and detailed definition by Earl and Katz (2006, p.45) is as follows:

[Data literacy is] a thinking process … of:
- standing back and deciding what you need to know and why
- collecting or locating the necessary data
- finding ways to link key data sources
- ensuring that the data are worth considering
- being aware of their limitations
- thinking about what the results mean … [and]
- Systematically considering an issue from a range of perspectives so that you really feel that you have evidence to explain, support and also challenge your point of view.

More recently, Riel, et. al. (2012) represents it in terms of digital literacies, shown in Table 1.

**Table 1 Classification of Digital Literacies (Riel, Christian, & Hinson, 2012)**

<table>
<thead>
<tr>
<th>Tools &amp; Interface</th>
<th>Information &amp; Data</th>
<th>Sharing &amp; Creation</th>
<th>Historical &amp; Cultural Context</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ Computational Basics</td>
<td>✓ Representation</td>
<td>✓ Intuitive Thinking</td>
<td>✓ Digital Citizenship</td>
</tr>
<tr>
<td>✓ Computer Hardware</td>
<td>✓ Search</td>
<td>✓ Documents(text)</td>
<td>✓ Diversity</td>
</tr>
<tr>
<td>✓ Computer Software &amp; Application</td>
<td>✓ Assembly</td>
<td>✓ Multimedia</td>
<td>✓ Intellectual Property</td>
</tr>
<tr>
<td>✓ Networks</td>
<td>✓ Analysis &amp; Judgement</td>
<td>✓ Communication</td>
<td>✓ Privacy and Identity</td>
</tr>
<tr>
<td>✓ Design</td>
<td>✓ Synthesis</td>
<td>✓ Online Persona</td>
<td>✓ Programmed Agendas</td>
</tr>
<tr>
<td>✓ Augmentation</td>
<td>✓ Archiving</td>
<td>✓ Productivity</td>
<td>✓ Technology Impact</td>
</tr>
<tr>
<td></td>
<td>✓ Navigation</td>
<td>✓ Sharing &amp; Collaboration</td>
<td></td>
</tr>
</tbody>
</table>

Koltay’s (2015) definition has a close connection with information literacy: ‘as a specific skill set and knowledge base, which empowers individuals to transform data into information and into actionable knowledge by enabling them to access, interpret, critically assess, manage, and ethically use data’. A more contemporary and concise synthesis is provided by Ridsdale et al., (2015, p.2): ‘Data literacy is the ability to collect, manage, evaluate, and apply data, in a critical manner.’ This is further described through the Table 2.

While we have defined *data literacy* as a component skill of being able to *discriminate* (Mason, Khan, & Smith, 2016), we now discern that a critical element is missing in all these definitions: *data production*. We are now living in an era of big data that would not exist if we were not all contributing to it. In other words, we need to better understand our place in the whole data lifecycle – we are both producers and consumers of data. This same point has been well made by Mirra, Morrell, and Filipiak, (2018) in their work on integrating critical theory with a theory of multiliteracies. It is also a pivotal perspective for how we conceive of curriculum interventions that are now needed.
A CURRICULUM PERSPECTIVE

In 1894, Herbert Spencer asked: ‘What knowledge is of most worth?’ For him, the answer was simple: ‘For – direct self-preservation – Science; for gaining livelihood – Science; for parental functions – Science; for good citizenship – Science; for the enjoyment of art – Science; for purpose of discipline – Science. Science is the best preparation for all these orders of activity.’ Modern curriculum is built on the same paradigm as envisaged by Spencer, but in our post-truth era, this needs an update. Moreover, the ‘4th paradigm’ of science – data intensive research – is now well underway (Hey, Tansley, & Tolle, 2009).

The digital world brings sweeping changes in our understanding, causing us to question our assumptions in every field. This is a transformative age where all relations, predictions, control and order are gradually becoming asymmetrical and fuzzy. All relations such as student and teacher, parent and child, subject and object, real or virtual, are changing their commonly understood shape and meanings.

It is widely accepted that many parents are concerned about the impact of social media on children and are interested in being involved with teachers to determine how their children understand, make sense of, and be safe in the digital world. Further education, until recently, only took account of social media or other internet platforms through media or social studies or as part of school-wide counselling programs that tended to focus only on online bullying or virtual safety. Of course, safety must be paramount. But it is not the only consideration when students are at the school level, particularly teenagers, as they are among the most prolific users of the Internet and social media. Teenagers, therefore, are the ones uniquely positioned to understand and evaluate the media’s role in their own lives and beyond.

There is now a universal concern over false news and fake data. This concern within the developed world mostly arises from the possibilities that through such news and data anyone can influence political, economic, and social well-being of a society. Recognizing this concern, researchers at MIT recently used a dataset of rumour cascades on Twitter from 2006 to 2017. The study found that about 126,000 rumours were spread by approximately three million people. The conclusion: false news reached more people than the truth; the top false news flows spread to between 1000 and 100,000 people, whereas the truth rarely dispersed to more than 1000 people and that falsehood also spread faster than the truth. The researchers also found that the degree of novelty and the emotional reactions that the news had on the readers may be a key responsible factor (Vosoughi et al., 2018).

Table 2 Dimensions of Data Literacy, following Ridsdale, et al., (2015)
The sixteenth century mathematician Cardano used to call the solutions to his algebraic equations that produced negative answers as *ficta* or fake (Ellenberg, 2015, p. 78). Now we accept negative or even imaginary numbers as essential to basic mathematical discourse. To deal with current fake data and artificial data visualizations that are reaching and panning out faster and through better means it is necessary that our future generations are better prepared and be skilled to deal with ‘the fake’. In the same way that negatives and imaginary numbers play a critical role in enhancing and extending human understanding to deal with real-life situations, teaching about fake data through data literacy presents a similar opportunity to potentially upgrade our cognitive abilities – from an answering and accepting mode to a questioning and rejecting mode; from *what* and *how*, to *why* first (Mason, 2012).

The need to develop critical skills with handling media is illustrated in a BBC game recently created that challenges young people to spot fake news through providing them with a virtual identity as an online reporter. The reporter must then mitigate challenges of false and fake news in the virtual world before publishing headline items. Similar websites such as Factitious test the news sense of visitors through mixed articles, some being fake and some real. It also provides hints to readers through links to original sources of the so-called news.

We suggest that data literacy be pursued within a reformed school curriculum through various ways, the most obvious being a subject focused on media and introduced at middle or senior school level to allow students to use social media to themes such as:

- Narratives on global change (e.g., global warming)
- Alternative identities and mixed realities
- Storytelling and data visualization
- Questioning why
- Personal and public data

*Narratives of Global Change*

Students can examine global warming using two perspectives – data evidence and risk evaluation. Risk evaluation and analysis can provide an organized, efficient and comparative way of looking at the form and structure of storytelling and its consequences on human action. Students may look at numerous claims and understand that binary oppositions play a role for the storytelling and faking data activity (Smith, 2012; Khan & Mason, 2016). Students may also discuss the book *Fake Science* (Ruse, 2017) which presents global warming as a scam.

*Alternative Identities*

Of course, we all have many identities throughout our lives – but the digital environment expands the range of options. Anonymity has played a pivotal role in the development of the Internet and platforms such as *Second Life* take role playing to a new level. Are we the same person in mixed realities? Do we have different ‘versions’ of ourselves? As technology develops, Case (2018) asks whether we will be able to tell the difference between a human and a cyborg?

*Storytelling and Data Visualization*

In broad historical terms, we have progressed from pre-literate societies where storytelling was the primary means of safeguarding cultural knowledge to an era where the ‘4th paradigm’ of science now needs ‘data storytellers’ as much as ‘data analysts’ (Khan & Mason, 2016). But as we all know, stories can be fictitious or factual.
**Questioning Why**
The digital environment provides unprecedented access to information; however, it is not yet an intelligent environment that scaffolds reasoning skills. To the contrary, with big data and the rise of ‘computer-aided explanation’ there are arguments that we may even need less reasoning skills (Mayer-Schönberger & Cukier, 2013; Nielsen, 2015). We question such conclusions and see them as colliding with the school curriculum that has embedded so-called ‘21st century skills’ in recent years. In an information-rich environment, in which storytelling features, cultivating reasoning and coherent rationale seems to be an essential learning skill.

**Personal and Public**
In the pre-digital economy, very few organizations collected personal and private data about citizens. In the new digital economy, however, data is the new currency. We voluntarily disclose it every time we subscribe to an Internet service. It is also often collected explicitly for the purpose of improving and customizing services to our needs. But in this process, we contribute to big data and help blur the boundary between personal and public data. Social networking sites specialize in sharing of personal information.

**AUSTRALIAN CURRICULUM PERSPECTIVES**
In 2012, the National School Improvement Tool was endorsed by all Australian state governments. It recognizes and highlights key domains of practice for school improvement including Analysis and Discussion of Data:

> A high priority is given to the school-wide analysis and discussion of systematically collected data on student outcomes, including academic, attendance and behavioural outcomes, and student wellbeing. Data analyses consider overall school performance as well as the performances of students from identified priority groups; evidence of improvement/regression over time; performances in comparison with similar schools; and, in the case of data from standardised tests, measures of growth across the years of school. (ACER, 2012, p. 4)

This statement was made in wake of Australia formalizing a national approach to the school curriculum. Since its first release in 2012 it has already been revised and it’s expected that the revision will continue as routine, even though the process is also heavily politicized. A related outcome is a national approach to standardized testing, known as NAPLAN (National assessment Program for Literacy and Numeracy).

As is evident from many other related documents (DoE, 2016; ACER, 2012) there is great emphasis on the data being collected through testing systems rather than the data students are involved with producing elsewhere on social media and related platforms such as Facebook Live, Google +, Instagram, Snapchat, LinkedIn and Twitter. Thus, there exists some disconnection between the production and analysis of data in developing an understanding of student capabilities in the digital age. Arguably, this disconnect can be understood as a deficiency in the educational system as well as an opportunity to broaden the scope of mix of essential capabilities to be inclusive of data literacy. The question arises: **how can the school curriculum address the need for data literacy skills of students in ways that can also broaden the scope of standardised testing?**
Apart from ‘ICT capability’ the Australian Curriculum has potential to address these skills through personal and social capabilities. These capabilities are defined by the curriculum to include a range of practices including recognizing and regulating emotions, developing empathy for others and understanding relationships, establishing and building positive relationships, making responsible decisions, working effectively in teams, handling challenging situations constructively and developing leadership skills. Educators may use this focus and use social media for teaching students to learn to understand themselves, others, and the broader life environment better. Including a richer definition of these capabilities might also lead to learning how to manage this learning more effectively. The curriculum also allocates several places where basic elements of data literacy may be addressed such as in Mathematics, Digital Technology, Social Science and Humanities. In each of these areas data literacy brings into focus essential questions such as:

- In what ways do we produce and use data in our daily lives?
- What influences shape the operation of a country’s news media and political system?
- How do we as citizens participate in an interconnected world?

The latter question listed above also aligns strongly with the ‘global citizenship education’ agenda advocated by UNESCO (2014) as a requirement to fulfilling the goals of the United Nations Secretary-General’s Global Education First Initiative (GEFI).

Of course, many other essential questions specific to key learning areas could be identified. As a pedagogical tool we may use the fake data paradigm to bring into focus some guiding questions in class such as: What are we seeing here? Why do we think this is so? What story is this data telling? What story is this data NOT telling and Why? Whose story is this? What are implications of this on the society? How can we find out more about this? How confident we are in our conclusions? Numerous case studies and examples for teaching purposes on area bias, social desirability bias, leading question bias, manipulation of averages, influencing through data visualization are discoverable (Ord, 2015; Best, 2004).

**CONCLUSION**

In our increasingly globalized world all citizens are increasingly involved in a broadening range of online social, commercial, and learning activities. In a ‘post-truth’ information economy this also involves creation, production and utilization of significant amounts of data. Because these activities can use highly personal data, it is critical that our students learn the skills to discriminate the increasingly blurred lines between fake and false, false and real, real and virtual, and virtual and imaginary.

This paper represents some initial probing of the issues that are becoming prominent following the proliferation of fake news and misinformation in the electronic media. Among our findings is that while we have increasing access to increasing volumes of data we are also exposed to a richer mix of storytelling. Data science does not supersede storytelling and so there is an urgent need to address data literacy in the school curriculum – not just as a skill for students to develop but also a perspective that educational systems need to consider in the calibration of high stakes standardised testing. Future educational planning and debate concerning schooling must also scrutinize the use of social media as an innovative teaching and learning tool, both inside and outside the classroom.
REFERENCES


