Hayashi, Y., et al. (Eds.) (2017). Workshop Proceedings of the 25th International Conference on Computers in Education. New Zealand: Asia-Pacific Society for Computers in Education

Learning to be Data Smart

Khalid KHAN * & Jon MASON
Charles Darwin University, Australia
*khalid.khan@cdu.edu.au

Abstract: Recent advances in the computing power have brought the processing of Big Data at the door step of an individual at a personal level. Pattern recognition, decision making, and modelling are some of the few skills that can be employed to make sense of personal data. In this short paper, we summarize some of our findings that highlight data literacy as a critical competency for 'smart learning'.

Keywords: data literacy, digital literacy, discriminate, correlations, smart learning.

1. Introduction

By 2030 it is predicted that automation, globalization and flexibility will change what we do in every job. With changes in the demands due to changes in job processes and advanced digital capabilities it is also predictable that society on average will spend considerable amount of time on learning skills *on* and *at* the job. This is not only a concern for the future; even current employers are looking for creativity and enterprise skills among their potential employees and they are ready to pay a premium to provide these skills. Job market research shows that the demand for critical thinking has increased by 158 percent in the last three years (FYA, 2017, p. 23). With algorithms and intelligent machines automating decision making processes, what key 'human' skills we need to carry to our jobs that can't be automated is the key question that is puzzling many minds.

We locate our current thinking on smart learning within the broader context of smart skills and the big data environment around us. Our positioning on data literacy goes beyond statistical learning and the construction of meaning from data through computational algorithms that try to make sense through mean, median, mode, deviation and related pattern finding mechanical procedures. We see the data, storyteller, and the context within which data is collected (or missed) as inseparable entities. Each of these entities is critical in the pursuit of sense making and discovery of knowledge hidden within data. Within this setting we consider data literacy as a form of mathematical literacy and critical thinking that is not confined and limited to the parameters of spreadsheets.

When words combine in a particular way, they make sentences that in turn combine in different ways to tell different stories, as do numbers and data. Stories, words and data combine, make sense. The woven stories engage and appeal to our imagination at a personal level. Combine these stories with visuals, graphs and colorful displays and we get engagement, emotions and a sense of meaning. Data literacy in this paradigm can be triangulated within the space of data, emotional literacy and mathematics/statistics. While emotional literacy can be understood as the ability to reflect on and exercise our own emotions, mathematics enables us to exercise our mind to draw conclusions objectively on the basis of pure logic and reasoning. It is therefore vital in the development of data literacy skills to make the link between data visualization and various mathematical representations and our emotional intelligence. Emotions distinguish humans from robots. Decision making for humans happens at both levels – emotional and reasoning. How many of our likes or dislikes are based purely on quantification?

The following discussion provides an overview of the issues drawn from our research over a period of last two years that focused on the topic of data literacy. For us, becoming *data literate* is about being *data smart*.

2. Data Literacy as a critical skill

The cognitive abilities required, in different jobs as listed by O*NET (a database that stores and updates information on skills required in different occupations sponsored by US Department of Labor), are:

- Deductive Reasoning
- Flexibility of Closure (ability to identify or detect a known pattern e.g., a figure, object, word, or sound, that is hidden in other distracting data)
- Speed of Closure (the ability to quickly make sense of…)
- Information Ordering
- Mathematical Reasoning
- Number Facility (ability to do basic operation on numbers quickly)
- Problem Sensitivity (involving seeing a hidden problem with in the problem)
- Selective Attention
- Visualization (the ability to imagine changes when parts are moved)

The *New Work Smarts* report (FYA, 2017, p. 22) provides data that show a critical low in some of the key skills required for work in the future. The report points out that lower level percentages that exists in Problem Solving, Digital Literacy and Mathematics proficiencies being around 35%, 27% and 45 % respectively. The figures are much higher for low socio-economic and Indigenous students. It is important to teach these skills within the curriculum or across curriculum during a student's school and educational life before the job instead of inculcating them at the job. We propose these skills and capabilities be considered as aspects of data literacy and taught cross-curriculum as a multi-disciplinary skill.

3. Black Box Artificial Intelligence, Algorithms and Mathematical Modelling

In order to avoid bias human decision-making more objective machine intelligence and algorithms have been employed. Machine intelligence is based on mathematical modeling. In a recent issue of MIT Technology Review Knight (2017) states that "Opaque and potentially biased mathematical models are remaking our lives". Likewise, a group of researchers researching social impacts of artificial intelligence has announced the *AI Now Initiative* in which one of the main research questions under study is Bias and Inclusion (Artificialintelligencenow.com, 2017):

> Data reflects the social and political conditions in which it is collected. AI is only able to "see" what is in the data it's given. This, along with many other factors, can lead to biased and unfair outcomes.

The bias in intelligent machines and algorithms has potentially negative consequences for disadvantaged communities and minorities. Even if the data is not influenced intentionally, the algorithms designed to predict patterns and correlation needs be carefully analyzed and not just believed to be correct.

The underlying assumptions based on which the systems are making their choices are not clear even to the systems' designers. It's not necessarily possible to determine which algorithms are biased and which ones are not (Spielkamp, 2017).

Algorithms are becoming ubiquitous on the web. The mathematical models and automated risk assessment that drive them are deciding who to call for job interviews and who to sanction in loan applications. Even judges are using these systems to decide on whether to grant bail applications (Center, 2017).

If the important decisions the algorithms make go unchecked, the financial and legal implications might be serious for equitable society and its social and cultural fabric. For a just society, it is vital that decision making is transparent and clear and not opaque.

In 2016, in a study by ProPublica – a nonprofit news organization that produces investigative journalism – conducted a study on risk scores on more than 7,000 persons arrested in a county in Florida between 2013 and 2014. ProPublica tested to find out how many of these people were charged with new crimes over the next two years, using the same weightings used by an algorithm (COMPAS) that is used by some judges in US court system. The result of the finding was that predictions for a repeat of violent crimes were only 20% in agreement with what actually happened. ProPublica's findings also highlighted 'significant racial disparities' (Julia A., 2017). They found:

> Black defendants were 77 percent more likely to be pegged as at higher risk of committing a future violent crime and 45 percent more likely to be predicted to commit a future crime of any kind.

The report also highlighted two key findings:

- The formula was particularly likely to falsely flag black defendants as future criminals, wrongly labeling them this way at almost twice the rate as white defendants.
- White defendants were mislabeled as low risk more often than black defendants.

The results for other similar risk analysis algorithms were found to be biased as well.

Legal systems have a long history of trying to predict of the chances of recommitting of crimes by the criminals about to be released. The racial factors such as race, nationality and skin color were often used to make such predictions. As late as late twentieth century this practice was common (Harcourt, 2016). These factors might have now seeped into machine algorithms and models design and the big question with big data algorithms, that follow no regulations is -*How do we know*? As technology progresses we soon cross a point in future where it would be impossible to explain the reasons how decision making happens within the algorithms and using AI may require – a leap of faith based on how smartly our human intuition and sense making are trained.

4. Smart Pedagogies and Australian Curriculum

With round-the-clock access to smart technologies our young generation is interacting with enormous amount of data these days. On one hand, they are recipients and consumers of data and information they can't make sense of and on the other hand, unknowingly and unwittingly, they act as a subject for the big data collection projects of corporations such as Facebook and Google. How do they make sense of this data, is a critical question. Many of the young adults neither have the tools nor being taught how to understand the data they are coming across or are part of.

> Smart learning has been defined as involving metacognitive aspects of learning: "It's not just what you know. It's what you know about what you know" (Paul, 2017).

Mathematics curriculums, presently, confines and limit data literacy to the teaching of statistical skills. They don't provide skills to students –framing of questions. What and which questions to ask, or recognize when data is presented in a misleading way or how the visualization of the data and the way it is graphed might have been manipulated. Students are required to be skilled to be skeptical and be better discriminators of information (Mason, Khan, and Smith, 2016).

Sense making happens when teachers allow lessons to be flexible, when they permit curiosity to take over their lessons. Teachers need to help students create and ask questions based on students' interaction with data. The National Council of Teachers of mathematics' Math Forum describes this aspect in the following words:

> The process of sense-making truly begins when we create questioning, curious classrooms full of students' own thoughts and ideas. By asking: What do you notice? What do you wonder? We give students opportunities to see problems in big-picture ways, and discover multiple strategies for tackling a problem. Self-confidence, reflective

skills, and engagement soar, and students discover that the goal is not to be "over and done," but to realize the many different ways to approach problems. (Mathforum.org, 2017)

Example 1

A content descriptor on data learning for year 3 and 4, within Australian Mathematics Curriculum describes that the student at this level: “Recognize different types of data and explore how the same data can be represented in different ways”. In the elaboration of the descriptor it has been explained under visual knowledge that student need to understand how visual elements create meaning (V7-5.australiancurriculum.edu.au, 2017). A food for thought for teachers is how to teach this and what activities to select to elaborate this particular aspect.

The *New York Times* (2017) recently acknowledged the need by creating two series within their learning network: *“What’s Going On in This Picture”* (WGOITPicture) and *“What Is Going On in This Graph”* (WGOITGraph). In the first weekly series, the *New York Times* invited teachers to discuss some of the pictures posted without any description within their class. The idea behind the series was asking students how they make sense of what they see when they look at an image, especially if that image comes with no caption, headline, links or other clues about its origins? Can constructing meaning from an image teach them something? Specifically, the following questions are what students can post their comments on: *What is going on in this picture? What do you see that makes you say that?*, and *What more can you find?*

Students are then supposed to post their remarks and read other students comments on *New York Times*. They are then able to participate with a facilitator teaching students ‘*visual thinking strategies’* by paraphrasing comments and linking to responses to help students’ understanding go deeper. At the end of the week the newspaper reveals the real information about the photo to help students understand, how the reading of the caption and story help people see the image differently. An example of a graph that can serve to discuss refugee crisis is the following:

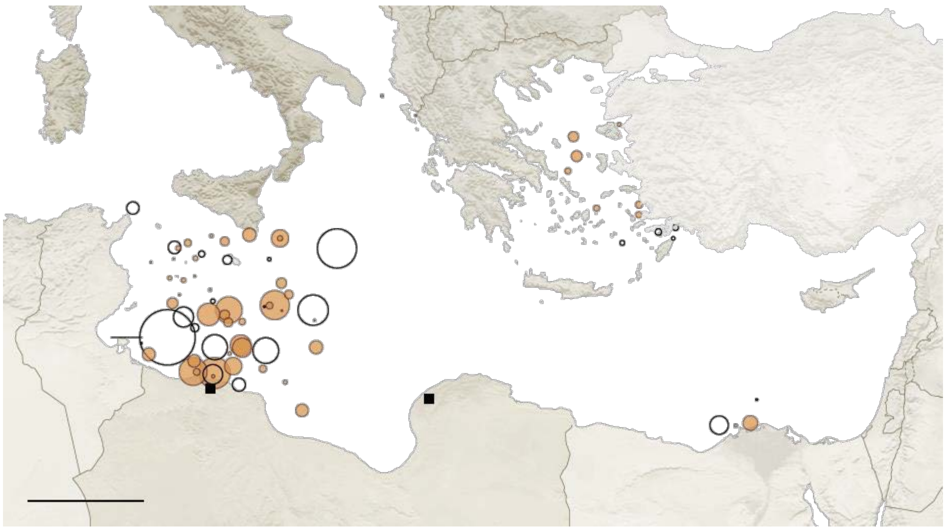

Figure 1. Source: (International Organization for Migration, 2017)

Each circle in above represents an incident, sized by the number of dead or missing within Mediterranean Sea around Libya. Without mentioning what the circles represent teachers may initiate a discussion on this world issue and then at a suitable stage explain the meaning of the circles. Unfilled circles are reports that have only been partly verified.

Example 2

Recently, in (Washington Post, 2017) in an editorial column it was reported that:

North Korean dictator Kim Jong-un] has shown no interest in talks — he won't even set foot in China, his biggest patron. Even if negotiations took place, the current regime has made clear that "*it will never place its self-defensive nuclear deterrence on the negotiating table*, as one envoy recently put it. [Emphasis added]

Jon Schwarz (2017), in a news article for *The Intercept*, reported what North Korea's Deputy UN Ambassador Kim In Ryong, actually had said:

> *As long as the U.S. hostile policy and nuclear threat continue* [emphasis added], the DPRK, no matter who may say what, will never place its self-defensive nuclear deterrence on the negotiation table or flinch an inch from the road chosen by itself, the road of bolstering up the state nuclear force.

This is a case of (intentionally) missing data to create a different story for the unprepared minds of readers. Due to the lack of skills in understanding the importance of missing data public can be manipulated to design and influence policies. Such examples can be taught in sociology or history classes to teach data literacy capabilities within the curriculum. An example that describes this in a mathematics class it may be explained is following –

Example 3

A teacher in year 11 or year 10 mathematics class asks students to simplify the following:

$$\sqrt[3]{7+5\sqrt{2}}+\sqrt[3]{7-5\sqrt{2}}=??$$

She asks one student to check the answer on WolframAlpha where the student finds the answer as follows:

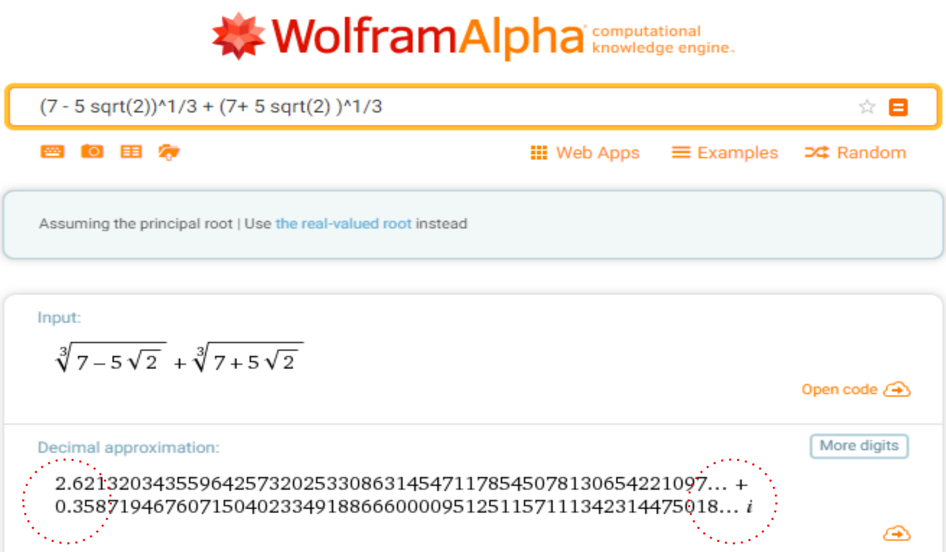

Figure 2. Problem solution by algorithms -WolframAlpha

However, she gave a completely different answer to the class via (all mathematically correct and verifiable) steps as follows:

$$let\ (7+5\sqrt{2})^{\frac{1}{3}} + (7-5\sqrt{2})^{\frac{1}{3}} = x$$

This Implies $[(7+5\sqrt{2})^{\frac{1}{3}} + (7-5\sqrt{2})^{\frac{1}{3}}]^3 = x^3$

Expanding brackets $[(7+5\sqrt{\cancel{2}}) + 3(7+5\sqrt{2})^{\frac{2}{3}}(7-5\sqrt{2})^{\frac{1}{3}} + 3(7+5\sqrt{2})^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{2}{3}} + (7-5\sqrt{\cancel{2}})] = x^3$

$$= 14 + 3[(7+5\sqrt{2})^{\frac{2}{3}}(7-5\sqrt{2})^{\frac{1}{3}} + (7+5\sqrt{2})^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{2}{3}}] = x^3$$

$$= 14 + 3[(7+5\sqrt{2})^{\frac{1}{3}}(7+5\sqrt{2})^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{1}{3}} + (7+5\sqrt{2})^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{1}{3}}] = x^3$$

$$= 14 + 3[(7+5\sqrt{2})^{\frac{1}{3}}[(7+5\sqrt{2})\ (7-5\sqrt{2})]^{\frac{1}{3}} + [(7+5\sqrt{2})\ (7-5\sqrt{2})]^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{1}{3}}] = x^3$$

$$= 14 + 3[(7+5\sqrt{2})^{\frac{1}{3}} \times (49-50)^{\frac{1}{3}} + (49-50)^{\frac{1}{3}}(7-5\sqrt{2})^{\frac{1}{3}}] = x^3$$

$$=14 + 3[(7+5\sqrt{2})^{\frac{1}{3}} \times (-1) + (-1)(7-5\sqrt{2})^{\frac{1}{3}}] = x^3$$

$$=14 - 3[(7+5\sqrt{2})^{\frac{1}{3}} + (7-5\sqrt{2})^{\frac{1}{3}}] = x^3$$

Thus $14 - 3x = x^3$

Which can be seen to possess $x = 2$ as the solution (check: 14 -3 x 2 = 14- 6 = 8)

Figure 3. Problem solution by teacher

A discussion on the design and assumptions on algorithms working behind the scenes may be initiated regarding the apparent and complete difference in the answers to the same question – *Why are the two processes (digital and manual) giving contradictory answers? What assumptions are made within mathematical paradigm of the digital technology?*

5. Position and summary of results

Research has shown that by improving a skills match to best practice can drive a 2% to 7% increase in the productivity in countries like Australia (OECD, 2015). The skills required were measures by OECD and found very closely linked to include written communication, maths, problem solving and digital literacy (OECD, 2015). Through our research (Khan, Mason, 2016; 2015; Mason et al., 2016) we found that in spite of rapid development and deployment of data analytics tools in recent years, there is a general lack and agreement on a common understanding on what skills are necessary for a data literate citizen and smart learning within the discourse on 21st century skills and competencies. The following is a summary of our positioning, and extends our previous list of points (Khan & Mason, 2016):

- As metaphor in reverse – data needs to be considered as *guilty until proven innocent.*
- As the giant Internet corporations take greater control of the entire data production and consumption lifecycle there is much at stake at a personal level.
- As a term, data *is* as much as data *are* – and academic pedantry will not change that;
- Data is not (necessarily) neutral.
- Data can be misused and misunderstood.
- There is an erroneous belief among data scientists that more data means more accurate predictions. It has been established time and again, with several examples that larger the data-higher the risk of error by coincidence due to of spurious correlations.

- Misrepresenting the patterns that may come by chance in one's data and thereby drawing *'false links'* is a big concern with algorithms dealing big data.
- Cultural and ethical dimensions need to be considered as key aspects of data literacy.
- Emergence of the era of *data-driven everything* presents new challenges for human sense-making.
- Story and the storyteller are contextually bound and cannot be separated;
- Asking key questions of the data is an art and science.
- Smart learning should shift the focus from *digital* to *data literacy*.
- Post-truth, the fake news era and big data analytics brings new realities in which any mix of data, information, and knowledge demands scrutiny and validation.
- What is missing from the data is at least as significant as what has been presented.
- Educators need to identify essential questions that require deep investigation both at cognitive and computational levels.
- To be *data smart* we need to create new and also refine the existing protocols for informed inquiry necessary in an age enabled and disrupted by digital innovation and ubiquitous data.
- *Data literacy* can be subsumed within a core skill of being discerning and *discriminate.*
- *Data literacy* is a form of Mathematical thinking that includes statistical literacy but not completely defined by it.
- Three literacies – information, data and statistical – are interrelated.
- Being *data smart* through learning skills in data *literacy* is missing from educational curricula.
- Teaching data literacy should involve combining, discriminating and aggregating different sources of data and in posing new questions and discovering new angles.
- As Big Data is moving from group predictions to individual predictions there are many unanswered questions. What happens to people's rights? Who owns *'my'* data? Is there any *my data*?
- What are the ethical dimensions of selling personal data to others without person's explicit consent?
- What is the future of 'smart decision making'? With advances in Artificial Intelligence 'how do we know decisions are fair and just'?

6. Conclusion

It is imperative that the new programs and educational frameworks are crafted to improve data literacy skills and recognize it being critically and fundamentally linked to the decimation of effective knowledge. The challenging part is to think creative ways and discover new and smart pedagogies that enable and make us *data smart*.

References

Artificialintelligencenow.com.(2017). *About: AI Now*. Available at: https://artificialintelligencenow.com/about [Accessed 19 Sep. 2017].

Center, E. (2017). *EPIC - Algorithms in the Criminal Justice System*. Epic.org. Available at: https://epic.org/algorithmic-transparency/crim-justice/ [Accessed 19 Sep. 2017].

FYA (2017). FYA| *The New Work Smarts Report.* Available at https://www.fya.org.au/report/ the-new-work-smarts/ [Accessed 2 Sep. 2017].

Harcourt, Bernard E. (2016*). Risk as a Proxy for Race.* (September 16, 2010). Criminology and Public Policy, Forthcoming; University of Chicago Law & Economics Olin Working Paper No. 535; University of Chicago Public Law Working Paper No. 323. Available at https://ssrn.com/abstract=1677654 [Accessed 19 Sep. 2017].

International Organization for Migration. (2017). *International Organization for Migration*. [online] Available at: https://www.iom.int/_[Accessed 12 Sep. 2017].

Julia, A.S. (2017). *Machine Bias —.* ProPublica. Available at:
https:// www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [Accessed 19 Sep. 2017].

Khan, K., & Mason, J. (2016). Data, the Story, the Storyteller, in Chen, W. et al. (Eds.) *Workshop Proceedings of the 24th International Conference on Computers*, India: Asia-Pacific Society for Computers in Education. (2016). pp. 142-144.

Khan, K. & Mason, J. (2015). Non-Numerical Aspects of School Mathematics. *Ogata, H. et al. (Eds.) Proceedings of the 23rd International Conference on Computers in Education. China: Asia-Pacific Society for Computers in Education. Dec15*.

Knight, W. (2017). *Biased algorithms are everywhere, and no one seems to care*. MIT Technology Review. Available at: https://www.technologyreview.com/s/608248/biased-algorithms-are-everywhere-and-no-one-seems-to-care/ [Accessed 19 Sep. 2017].

Mathforum.org. (2017). *The Math Forum.* Available at: http://mathforum.org/pow/noticewonder [Accessed 12 Sep. 2017].

Mason, J., Khan, K., &. Smith, S. (2016). Literate, Numerate, Discriminate – Realigning 21st Century Skills, in Chen, W. et al. (Eds.) *Proceedings of the 24th International Conference on Computers in Education*. India: Asia-Pacific Society for Computers in Education. pp. (2016) 609-614.

Nytimes.com. (2017). *What's Going On in This Picture?.* Available at: https://www.nytimes.com/column/learning-whats-going-on-in-this-picture_[Accessed 12 Sep. 2017].

OECD (2015). *Labour Market Mismatch and Labour Productivity Evidence from PIAAC Data*. Available at: https://www.oecd.org/eco/growth/Labour-Market-Mismatch-and-Labour-Productivity-Evidence-from-PIAAC-Data.pdf__[Accessed 2 Sep. 2017].

Paul, A. (2017). *Smart Learning Strategies*. The Creativity Post. Available at: http://www.creativitypost.com/education/smart_learning_strategies_[Accessed 19 Sep. 2017].

Schwarz, J. (2017). *North Korea Keeps Saying it Might Give Up its Nuclear Weapons — But Most News Outlets Won't Tell You That.* [online] The Intercept. Available at:
https://theintercept.com/2017/08/25/north-korea-keeps-saying-it-might-give-up-its-nuclear-weapons-but-most-news-outlets-wont-tell-you-that/_[Accessed 12 Sep. 2017].

Spielkamp, M. (2017). *Inspecting Algorithms for Bias*. [online] MIT Technology Review. Available at: https://www.technologyreview.com/s/607955/inspecting-algorithms-for-bias/ [Accessed 19 Sep. 2017].

V7-5.australiancurriculum.edu.au. (2017). *Digital Technologies Foundation to Year 10 Curriculum by rows - The Australian Curriculum v7.5*. Available at: http://v7-5.australiancurriculum.edu.au/technologies/digital-technologies/curriculum/f-10?layout=1#cdcod e=ACTDIK008&level=3-4 [Accessed 13 Sep. 2017].

Washington Post. (2017). Opinion | *Why aren't we talking about regime change in North Korea?* Available at: https://www.washingtonpost.com/opinions/global-opinions/why-arent-we-talking-about-regime-change-in-north-korea/2017/08/31/d36b0b00-8d9e-11e7-91d5-ab4e4bb76a3a_story.html [Accessed 12 Sep. 2017].