Drowning in the Data Deluge

George E. Andrews [(1)(2)]

Abstract

The Data Deluge and its digital enablers are a gargantuan phenomenon in science. There are many accompanying side effects. These range from the evaluations of scientific research and researchers to how we teach mathematics to children. This article takes a close look at these matters.

1 Introduction

In 2010, Sastry Pantula, director of the Division of Mathematical Sciences (DMS) at the National Science Foundation, proposed changing the DMS name to the Division of Mathematical and Statistical Sciences (DMSS). In response to this proposal Eric Friedlander, President of the American Mathematical Society (AMS) sent an e-mail message to each AMS member requesting feedback on this proposal. Attached to his message was a letter from Director Pantula giving his justification for the change.

I use this incident to begin to explain the object of this article, namely, the side effects of the Data Deluge. Pantula's letter provides a nice starting point. Here is the relevant paragraph from his letter:

> Big data provide big opportunities for mathematical and statistical sciences. It is an exciting time for our Division. In his FY12 budget roll-out speech NSF Director Dr. Subra Suresh referred to the "era of data and observation." The NSF 2011-2016 Strategic Plan notes that "The revolution in information and communication technologies is another major factor influencing the conduct of 21st century research. New cyber tools for collecting, analyzing, communicating, and

(1) Partially supported by NSA Grant H98230-12-1-0205.

(2) This is the written version of the author's American Mathematical Society Retiring Presidential Address delivered at the Joint AMS/MAA meeting in Boston on January 5, 2012.

> storing information are transforming the conduct of research and learning. One aspect of the information technology revolution is the 'DATA DELUGE,' shorthand for the emergence of massive amounts of data and the changing capacity of scientists and engineers to maintain and analyze it." Extracting useful knowledge from the deluge of data is critical to the scientific successes of the future. Data-intensive research will drive many of the major scientific breakthroughs in the coming decades. There is a long-term need for research and workforce development in computational and data-enabled sciences. Statistics is broadly recognized as a data-centric discipline, thus having it in the Division's name as proposed would be advantageous whenever "Big Data" and data-sciences investments are discussed internally and externally.

In the next section, I will examine the current Data Deluge enthusiasm. This will lead naturally to "metrics," the topic for Section 3. All of this then fits into the thought of Marshall McLuhan on "The Medium is the Message."

I conclude with my concerns about the negative effects of the Data Deluge on mathematics education and in particular, on the Common Core State Standards for Mathematics. This makes grass roots professional development projects all the more important.

2 The Data Deluge

There is remarkable restraint in the paragraph by Pantula that I quoted in Section 1. To understand the widespread giddy enthusiasm for the Data Deluge we turn to an article by Chris Anderson in Wired Magazine [2]. A few quotes from his article clearly suggest that we are at the cusp of a scientific "paradigm shift" in the sense of Thomas Kuhn [12]:

> "All models are wrong, but some are useful."
>
> So proclaimed statistician George Box 30 years ago...
>
> Peter Norvig, Google's research director, offered an update to George Box's maxim: "All models are wrong, and increasingly you can succeed without them."
>
> This is a world where massive amounts of data and applied mathematics replace every other tool that might be brought to bear....With enough data, the numbers speak for themselves...
>
> The scientific method is built around testable hypotheses. These models, for the most part, are systems visualized in the minds of scientists.

> The models are then tested, and experiments confirm or falsify theoretical models of how the world works. This is the way science has worked for hundreds of years. Scientists are trained to recognize that correlation is not causation, that no conclusions should be drawn simply on the basis of correlation between X and Y (it could just be a coincidence). Instead, you must understand the underlying mechanisms that connect the two. Once you have a model, you can connect the data sets with confidence. Data without a model is just noise. But faced with massive data, this approach to science – hypothesize, model, test – is becoming obsolete. Consider physics: Newtonian models were crude approximations of the truth (wrong at the atomic level, but still useful). A hundred years ago, statistically based quantum mechanics offered a better picture but quantum mechanics is yet another model, and as such it, too, is flawed, no doubt a caricature of a more complex underlying reality. The reason physics has drifted into theoretical speculation about n-dimensional grand unified models over the past few decades (the "beautiful story" phase of a discipline starved of data) is that we don't know how to run the experiments that would falsify the hypotheses – the energies are too high, the accelerators too expensive, and so on.
>
> Now biology is heading in the same direction....
>
> In short, the more we learn about biology, the further we find ourselves from a model that can explain it.
>
> There is now a better way. Petabytes allow us to say: "Correlation is enough." We can stop looking for models. We can analyze the data without hypotheses about what it might show. We can throw the numbers into the biggest computing clusters the world has ever seen and let statistical algorithms find patterns where science cannot.

The article then provides an account of the amazing achievements of J. Craig Venter among them the genetic sequencing of the Sargasso Sea [1]. The article concludes:

> Learning to use a "computer" of this scale may be challenging. But the opportunity is great: The new availability of huge amounts of data, along with the statistical tools to crunch these numbers, offers a whole new way of understanding the world. Correlation supersedes causation, and science can advance even without coherent models, unified theories, or really any mechanistic explanation at all.
>
> There's no reason to cling to our old ways. It's time to ask: What can science learn from Google?

If this prospect is even close to an accurate picture of the future, then it is no wonder that the Data Deluge has generated immense excitement. Clearly

Anderson's description portends the greatest paradigm shift in the history of science. After all, the Copernican revolution is restricted to astronomy, and Pasteur's germ theory affected medicine but not physics. However, if the Data Deluge disposes of experiment and theory in science, it affects everything from astronomy to zoology.

On the other hand, Wired Magazine is a popular magazine. So let us turn to a more conventional institution for its take. The National Science Foundation maintains a website [9] devoted to the program: Computational and Data-Enabled Science and Engineering (CDS&E). Here are relevant statements from that website:

> Computational and Data-Enabled Science and Engineering (CDS&E) is a new program. CDS&E is now clearly recognizable as a distinct intellectual and technological discipline lying at the intersection of applied mathematics, statistics, computer science, core science and engineering disciplines....
>
> We regard CDS&E as explicitly recognizing the importance of data-enabled, data-intensive, and data centric science. CDS&E broadly interpreted now affects virtually every area of science and technology, revolutionizing the way science and engineering are done. Theory and experimentation have for centuries been regarded as two fundamental pillars of science. It is now widely recognized that computational and data-enabled science forms a critical third pillar....
>
> NSF can make a strong statement that will lead the Foundation, researchers it funds, and US universities and colleges generally, by recognizing CDS&E as the distinct discipline it has clearly become.

What is the difference in viewpoint between the NSF statements and the Anderson article in Wired Magazine? Both Wired and the NSF see CDS&E as a "critical third pillar" in scientific research. The other two pillars, experiment and theory, are viewed by both as ancient. The NSF is not ready to discard experiment and theory. Wired sees them as yesterday's methodology.

It becomes clear that the Data Deluge is the current Wave of the Future or, at least, is so regarded by many. The problem is that when "waves of the future" show up they often wash away a number of worthy things and leave a number of questionable items littering the beach.

I do not dispute the immensity of the Data Deluge and its great potential for substantial advances in many branches of science. I do, however, hope that a sense of proportion can be maintained.

I fear that one of the accompanying, unintended consequences is the unstated assumption that *nothing* is trustworthy if it is not supported by *data*. This then raises the question of how one applies data in the evaluation of academic activities, and this leads to metrics.

3 Metrics

The most ambitious effort to use metrics extensively is STAR METRICS (Science and Technology for America's Reinvestment- Measuring the EffecT of Research on Innovation, Competitiveness and Science) [18].

> STAR METRICS is a federal and research institution collaboration to create a repository of data and tools that will be useful to assess the impact of federal R&D investments. The National Institutes of Health (NIH) and the National Science Foundation (NSF), under the auspices of Office of Science and Technology Policy (OSTP), are leading this project....The STAR METRICS project consists of two implementation levels:
>
> Level I: Developing uniform, audit-able and standardized measures of the impact of science spending (ARRA and non-ARRA) on job creation, using data from research institutions existing database records. No personally identifiable information (PII) is collected in Level I.
>
> Level II: Developing measures of the impact of federal science investment on scientific knowledge (using metrics such as publications and citations), social outcomes (e.g. health outcomes measures and environmental impact factors), workforce outcomes (e.g. student mobility and employment), and economic growth (e.g. tracing patents, new company start-ups and other measures). Data elements that will be collected in Level II will be collectively determined in consultation with Institutions that have joined Level I.

In a slide presentation [4], the entire project is summarized thus on slide 9:

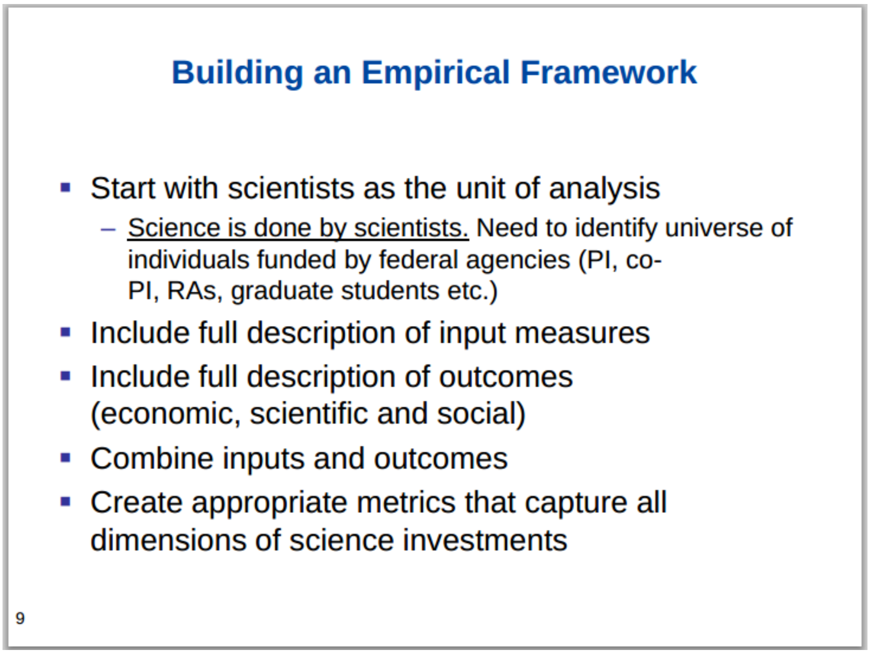

Figure 1: STAR METRICS presentation summary slide

The final point makes clear the all encompassing goal of STAR METRICS:

> *Create appropriate metrics that capture all dimensions of science investments.*

There are many metrics already in use. Let us look at a few. Perhaps the metric most widely known in the mathematics community is the Impact Factor. Its notoriety is probably a result of the article, Nefarious Numbers [3] by D. Arnold and K. Fowler. They start with the definition:

> The impact factor for a journal in a given year is calculated by ISI (Thomson Reuters) as the average number of citations in that year to the articles the journal published in the preceding two years.

They then cite a number of criticisms:

- A journal's distribution of citations does not determine its quality.
- The impact factor is a crude statistic, reporting only one particular item of information from the citation distribution.

- It is a flawed statistic. For one thing, the distribution of citations among papers is highly skewed, so the mean for the journal tends to be misleading.

  For another, the impact factor only refers to citations within the first two years after publication (a particularly serious deficiency for mathematics, in which around 90% of citations occur after two years).
- The underlying database is flawed, containing errors and including a biased selection of journals.
- Many confounding factors are ignored, for example, article type (editorials, reviews, and letters versus original research articles), multiple authorship, self-citation, language of publication, etc.

Next they make a statement which is central to all of the worries about metrics:

> Despite these difficulties, the allure of the impact factor as a single, readily available number—not requiring complex judgments or expert input, but purporting to represent journal quality—has proven irresistible to many.

This last quote is applicable to many metrics. The bulk of their article is devoted to manipulation of the impact factor.

It is generally acknowledged that the impact factor is not a good measure of individuals. To fill this gap, J. Hirsch [11] invented the h-index:

> "A scientist has index h if h of his or her Np papers have at least h citations each and the other $(Np - h)$ papers have h citations each."

He also defines the m-index as h/n where n is the number of years since the individual's first published paper.

> Based on typical h and m values found, I suggest (with large error bars) that for faculty at major research universities, $h \approx 12$ might be a typical value for advancement to tenure (associate professor) and that $h \approx 18$ might be a typical value for advancement to full professor. Fellowship in the American Physical Society might occur typically for $h \approx 15 - 20$. Membership in the National Academy of Sciences of the United States of America may typically be associated with $h \approx 45$ and higher, except in exceptional circumstances.

Of course, the objections to the impact factor raised by Arnold and Fowler can be lodged against the h-index with only minor alterations.

Proponents of the various metrics are not unaware of these criticisms. In response, numerous new metrics have been produced to answer these objections (e.g. the c-index, $h2$ lower, $h2$ central, $h2$ upper, etc.[5],[6]).

The loss of information involved in this type of "bean counting" is beautifully illustrated by considering two papers by the late Henry Mann:

> H.B. Mann & D.R. Whitney, On a test of whether one of two random variables is stochastically larger than the other, Ann. Math. Stat. 18(1947), 50–60.
>
> H.B. Mann, A proof of the fundamental theorem on the density of sums of sets of positive integers, Ann. of Math., 43(1942), 523–527.

The first paper has 2067 citations according to the Web of Knowledge and the second has 28. The second paper won the 1946 Cole Prize in Number Theory [17].

The first paper provides the moments of the distribution connected with nonparametric statistics, and the second solved a major open problem in number theory. These two papers are very important in *incomparably* different ways. A reduction of their importance to any sort of citation count is absurd.

Next we look at the work of Fields medalists and a table prepared by Penn State Science Librarian, John J. Meier [16]:

Citations of most cited work	Number of Medalists
500+	4
400–499	8
300–399	10
200–299	9
100–199	6
50–99	9
1–49	4

Table 1: Citations for Fields Medalists

If a citation count was a good measure of the importance of one's work, we would expect that the vast majority of Field's Medalists would be concentrated at the top of this table. Of course, if one believes that citation counts are the gold standard for measuring quality, then one must conclude that at least 19 (or perhaps 28) Medalists really did not deserve the Fields Medal.

Almost every mathematics professor is required to collect student evaluations. These come in two flavors: (1) mark a box from 1, horrible, to 10, outstanding, or (2) write comments. The first of these provides, again, a convenient metric.

Here are two examples of written student evaluations of the same professor (whose identity I have promised to keep secret) taken from his large lecture class:

1. What this class needs is free beer, dancing girls, and pot.

2. The consistent quality of Professor X's communication skills, thoroughness, clarity, anticipation of likely student problems, and helpful attitude make him a *superior* instructor. In a rarity for Penn State, he stressed the derivation of concepts to deepen the understanding of their use instead of struggling through a proof without stating its relevance and then saying "Just use the formula."

It is clear that reducing these two comments to numbers loses an immense amount of information both about the professor *and* the students.

4 "The Medium is the Message"

My greatest concern with the Data Deluge is the all encompassing role played by data. As I said earlier, there is a ubiquitous, unstated assumption that *nothing* is trustworthy if it is not supported by data. Indeed, data is the new medium for communicating a variety of evaluations.

This leads us to the controversial, flamboyant, enigmatic media guru of the 1960's and '70's, Marshall McLuhan. It was he who popularized the phrase "The Medium is the Message." This is akin to a modernized Hegelian zeitgeist.

McLuhan begins his book, Understanding Media, with an explanation of this phrase [15, p.7]

> In a culture like ours, long accustomed to splitting and dividing all things as a means of control, it is sometimes a bit of a shock to be reminded that, in operational and practical fact, the medium is the message. This is merely to say that the personal and social consequences of any medium—that is, of any extension of ourselves—result from the new scale that is introduced into our affairs by each extension of ourselves, or by any new technology.

McLuhan emphasizes this point again and again [15, p.11]:

> ...a few years ago, General David Sarnoff made this statement: 'We are too prone to make technological instruments the scapegoats for the sins of those who wield them. The products of modern science are not in

> themselves good or bad: it is the way they are used that determines their value.'
>
> That is the voice of the current somnambulism.

and [15, p.18]

> Our conventional response to all media, namely that it is how they are used that counts, is the numb stance of the technological idiot.
>
> For the 'content' of the medium is like the juicy piece of meat carried by the burglar to distract the watchdog of the mind.

Mark Federman [10] concludes his article on "the medium is the message" with the following paragraph:

> Why is this understanding of "the medium is the message" particularly useful? We tend to notice changes—even slight changes (that unfortunately we often tend to discount in significance). "The medium is the message" tells us that noticing change in our societal or cultural ground conditions indicates the presence of a new message, that is, the effects of a new medium. With this early warning, we can set out to characterize and identify the new medium before it becomes obvious to everyone—a process that often takes years or even decades. And if we discover that the new medium brings along effects that might be detrimental to our society or culture, we have the opportunity to influence the development and evolution of the new innovation before the effects become pervasive. As McLuhan reminds us, "Control over change would seem to consist in moving not with it but ahead of it. Anticipation gives the power to deflect and control force." [15, p.199]

I wish to emphasize the central importance of the fact that a medium seeks content that is appropriate to it, and it ignores content that it cannot easily accommodate. Metrics of all sorts are very much the type of instruments naturally required in the medium of data for comparisons of large data sets. Indeed, once we have placed our faith uniquely in data supported evidence, we can naturally expect that decision making will often become automated. This is the point made in the May, 2011 McKinsey Report on Big Data [14]. In a section entitled: *Replacing/Supporting Human Decision Making With Automated Algorithms*, we find

> Sophisticated analytics can substantially improve decision making, minimize risks, and unearth valuable insights that would otherwise remain hidden... . Decision making may never be the same ...

The role of data in the decision making of the academy merits special attention. It is especially disturbing to see this new medium of evaluation settle into place almost unnoticed.

In the final sections of this paper, I will concentrate on general educational issues that are being inadvertently shaped by the Data Deluge and its accompanying digital tools.

5 Education and the CCSSM

Education is a continuing concern for the future of our country. My goal here is to consider recent efforts for improvement and to examine the ways in which these efforts have been undermined by inadequate attention to the Data Deluge and its friends.

The Common Core State Standards for Mathematics (CCSSM)[8] and those for English (CCSSE) [7] were released in 2010. I will mostly discuss CCSSM with only one reference to CCSSE.

The CCSSM was prepared by a large committee (chaired by William McCallum) which was created by the National Governors Association and the Council of Chief State School Officers. it is a coherent and mathematically sound set of standards, and the AMS Committee on Education has rightly given a firm endorsement.

Unfortunately coherence and mathematical soundness alone are not adequate to produce a successful implementation of these standards. I am especially concerned about the emerging role of Big Data and its unnoticed role in the CCSSM. A number of omissions or near omissions suggest that an over-concern with data, statistics and probability may produce curricula that tend to ignore how real human beings actually learn things.

To gain a hint of what is regarded as important let us see some of the words used often. We start with *big data* friendly words (the irony of using data here is noted).

WORD	APPEARANCES IN CCSSM
Data	159
Probability	97
Statistics (and variants)	52
Technology	17
Computer	10
Spreadsheet	7

Table 2: Big Data friendly word frequencies in CCSSM

But now, what about words related to classical K-12 mathematics and the way real human beings learn?

WORD	APPEARANCES IN CCSSM
Geometry/Geometric	118
Algebra (and variants)	75
Arithmetic	27
Memory	2
Mnemonic	2 [a]
Memorization	1 [b]
Pencil	1
Freehand	1
By hand	1
Rote	0

[a] in one sentence on FOIL
[b] in a reference title

Table 3: Classical K-12 mathematics word frequencies in CCSSM

The first thing to jump out of these word counts is the fact that "data" overwhelms all the other words. Secondly we note that the geometry-algebra-arithmetic triad occurs 220 times, while the data-probability-statistics triad occurs 308 times. Let us dig deeper into the appearances of the final seven words in the second list.

Memory occurs first:

> In grade 2: Fluently add and subtract within 20 using mental strategies. By end of grade 2, know from memory all sums of two one-digit numbers.

And finally:

> In grade 3: Fluently multiply and divide within 100, using strategies such as the relationship between multiplication and division (e.g. knowing that $8 \times 5 = 40$, one knows $40/5 = 8$) or properties of operations. By the end of grade 3, know from memory all products of two one-digit numbers.

Surely students need to know many more things from memory. Nonetheless the CCSSM does not prescribe anything either geometric or algebraic that should be embedded in memory.

The only time a mnemonic is mentioned is in a discussion of the uniquely pernicious mnemonic FOIL (first, outer, inner, last) from page 4:

> There is a world of difference between a student who can summon a mnemonic device to expand a product such as $(a+b)(x+y)$ and a student who can explain where the mnemonic comes from. The student who can explain the rule understand the mathematics, and may have a better chance to succeed at a less familiar task such as expanding $(a+b+c)(x+y)$.

The implication is clear. Stay away from mnemonics, and mnemonic is as close as the text ever gets to memorization, or rote learning.

Contrast this mindset with the words of an Illinois High School teacher in a letter to me:

> Memorization for its own sake is admittedly of limited value: however, anyone who has learned mathematics in a rigorous manner attests to the fact that post-comprehension memorization is beneficial to promote efficiency in problem-solving. Our reform advocates over the past 20 or 25 years unfortunately have been permitted to equate in the mind of educators memorization with tedium and lack of understanding; its as if quick command of the facts and comprehension were somehow mutually exclusive.

Finally we come to one of the biggest difficulties with the Common Core in both mathematics and English. This problem is epitomized by the one appearance of "pencil" in the CCSSM [8, p.7]:

> Mathematically proficient students consider the available tools when solving a mathematical problem. These tools might include pencil and paper, concrete models, a ruler, a protractor, a calculator, a spreadsheet, a computer algebra system, a statistical package, or dynamic geometry software.

At least "concrete models, a ruler, a protractor" weren't bundled into "static geometry hardware."

If we put this de-emphasis of "pencil and paper" into a wider context, our concerns come into sharp focus. Here is the guide for writing for K–2 from the Common Core Standards for English Language Arts and Literacy [7, p.19].

> With guidance and support from adults, explore ("use" replaces "explore" in grades 1 and 2) a variety of digital tools to produce and publish writing, including in collaboration with peers.

Why is this reliance on digital tools without mention of handwriting troublesome? There is evidence that we are ignoring how real human beings actually learn.

A. Mangen and J.-L. Velay [13] present a lengthy account of why the physical act of handwriting is important. Here is a list of their discoveries from *Writing by hand better for learning, study shows* [19].

> Don't write off pen and paper just yet. New research shows the old-fashioned tools can make you a stronger learner.
>
> Sure, for many, writing by hand seems a little retro. However, using a keyboard or touchscreen to write is a drastically different cognitive process from writing by hand, according to a study published in the journal Advances in Haptics.
>
> Researchers Anne Mangen, of the University of Stavanger in Norway, and Jean-Luc Velay, a French neuroscientist, said their research indicates the increase in digital writing in schools needs to be examined more closely....
>
> Penmanship is not required to be taught in Illinois schools. And cursive is not part of the national Common Core State Standards Initiative, which is adopted by most U.S. states, while keyboarding skills are specifically required.
>
> The researchers found that writing by hand is fundamentally different from typing on a computer. And people who are learning new letters—such as children learning to read for the first time, or as adults picking up a second language with new characters—retain the information best when writing the letters by hand, according to Velay's research.
>
> The physical act of holding a pencil and shaping letters sends feedback signals to the brain. This leaves a "motor memory," which later makes it easier to recall the information connected with the movement, according to the study....
>
> The movement for "the typing of a 'T' is no different than the typing of a 'Y'," Mangen said. Further, when "you write something on the keyboard, you get the visual output somewhere else, on the screen," as opposed to you watching your hand when you write on paper, she said.
>
> This means that learning to write by hand can strengthen reading skills.
>
> Mangen said she understands the benefits of typing—it's quite simply faster. However, the fact that writing by hand can be comparitively "long and difficult" might be the reason it can be so helpful to triggering brain processes, she said.

Everything discussed in the article just quoted makes eminent common sense. Indeed it is additionally disturbing to think that we need another "study" in order to make a case that is obvious to any experienced teacher who is paying attention.

We are no longer able to assert merely that "Grass in green!" Instead we must add something like:

> A team of Harvard scientists has studied 9328 blades of grass from 37 randomly selected countries. They measured the wave length of light emanating from each blade when placed in the noonday sun in Harvard Yard. 98.32% produced light of wave length between 520 and 570 nanometers which is the accepted standard measure for green as certified by the International Bureau of Standards.

How convinced would you be if I had substituted for the Mangen-Velay study the following quote (supplied by Richard Escobales) of the late, great mathematician S.S. Chern:

> My mathematical strength lies in my ability in computation. Even now I do not mind doing lengthy computations, while years ago I could do them with relatively few errors. This is a training which is now relatively unpopular and has not been encouraged. It is still a great advantage in dealing with many problems.

Would we really be worse off if we relied more on the wisdom and common sense of the eminently successful and less on "studies?"

In any event, common sense is most likely to prevail in education when it is practiced in the classroom by experienced, knowledgeable teachers, and this leads to the next and final section.

6 Professional Development

Up to this point, I have described a world enthralled with data and the digital machines that enable it. I have pointed out the ways in which this narrowing of our vision obscures important information, distorts evaluations, and impairs learning.

In this final section, I want to encourage efforts that might at least mitigate some of the woes of mathematics education is the US today.

A big problem with the Common Core State Standards is that they are *top down* rather than bottom up. A serious effort at the professional development of teachers is required if these new standards have any chance of improving education. Without confident, knowledgeable teachers in the classroom, improvement in mathematics education is almost certainly impossible.

There are a number of projects undertaken by mathematics faculty and mathematics departments which directly involve classroom teachers with the object of (1) increasing their knowledge of relevant mathematics and (2) helping them communicate this knowledge to their students.

If there is any hope of directing education so that the insidious aspects of a digital and data suffused world can be mitigated, professional development of teachers' content knowledge is certainly a good first step.

Here is a very incomplete list of active people working on relevant projects. In each case, I list at least one project or book along with a web address.

Scott Baldridge, Louisiana State University, The Baker School Project, https://www.math.lsu.edu/~sbaldrid/

Deborah Ball, University of Michigan, Center for Proficiency in Teaching Mathematics, http://www-personal.umich.edu/~dball/

Hy Bass, University of Michigan, Center for Proficiency in Teaching Mathematics, http://www.soe.umich.edu/people/profile/hyman_bass/

Amy Cohen, Rutgers University, New Jersey Partnership for Excellence in Middle School Mathematics, http://www.math.rutgers.edu/~acc/

Ken Gross, University of Vermont, Vermont Mathematics Initiative, http://www.cems.uvm.edu/~gross/

Jim Lewis, University of Nebraska, NebraskaMATH, http://www.math.unl.edu/~wlewis1/

Tom Parker, Michigan State University, Books (with Scott Baldridge): Elementary Mathematics for Teachers, Elementary Geometry for Teachers, http://www.math.msu.edu/~parker/

Hung-Hsi Wu, University of California-Berkeley, Book: Understanding Numbers in Elementary School Mathematics, http://math.berkeley.edu/~wu/

I would urge everyone who has read this far to consider how you might get actively involved with these or similar projects.

References

[1] J.C. Venter et al. "Environmental genome shotgun sequencing of the Sargasso Sea". In: *Science* 304 (2004), pp. 66–74.

[2] Chris Anderson. "The end of theory: the data deluge makes the scientific method obsolete". In: *Wired Magazine* 16.07 (June 23, 2008). URL: http://www.wired.com/science/discoveries/magazine/16-07/pb_theory.

[3] D. Arnold and K. Fowler. "Nefarious numbers". In: *Notices of the Amer. Math. Soc.* (58 2011), pp. 434–437.

[4] S. Bertuzzi. *STAR METRICS*. URL: http://www.vivoweb.org/files/StarMetrics.pdf.

[5] Lutz Bornmann, Rdiger Mutz, and Hans-Dieter Daniel. "The h index research output measurement: Two approaches to enhance its accuracy". In: *Journal of Informetrics* 4.3 (2010), pp. 407 –414. URL: http://www.sciencedirect.com/science/article/pii/S1751157710000271.

[6] Maria Bras-Amors, Josep Domingo-Ferrer, and Vicen Torra. "A bibliometric index based on the collaboration distance between cited and citing authors". In: *Journal of Informetrics* 5.2 (2011), pp. 248 –264. ISSN: 1751-1577. DOI: 10.1016/j.joi.2010.11.001. URL: http://www.sciencedirect.com/science/article/pii/S1751157710000970.

[7] *Common Core State Standards for English Language Arts & Literacy in History/Social Studies, Science and Technical Subjects*. URL: http://www.corestandards.org/assets/CCSSI_ELA/%20Standards.pdf.

[8] *Common Core State Standards for Mathematics*. URL: http://www.corestandards.org/assets/CCSSI_Math/%20Standards.pdf.

[9] *Computational and Data-Enabled Science and Engineering*. URL: http://www.nsf.gov/mps/cds-e/.

[10] Mark Federman. *What is the Meaning of the Medium is the Message?* July 23, 2004. URL: http://individual.utoronto.ca/markfederman/article_mediumisthemessage.htm (visited on 01/16/2012).

[11] J. E. Hirsch. "An index to quantify an individual's scientific research output". In: *Proceedings of the National Academy of Sciences of the United States of America* 102.46 (2005), pp. 16569–16572. DOI: 10.1073/pnas.0507655102. eprint: http://www.pnas.org/content/102/46/16569.full.pdf+html. URL: http://www.pnas.org/content/102/46/16569.abstract.

[12] T. S. Kuhn. *The Structure of Scientific Revolutions*. 3rd. Chicago: Univ of Chicago Press, 1996.

[13] A. Mangen and J.-L. Velay. "Digitizing literacy: reflections on the haptics of writing". In: *Advances in Haptics*. Ed. by M. H. Zadeh. Rijeka, Croatia: InTech, 2010. Chap. 20, pp. 385–401.

[14] James Manyika et al. *Big data: The next frontier for innovation, competition, and productivity*. 2011. URL: http://www.mckinsey.com/Insights/MGI/Research/Technology_and_Innovation/Big_data_The_next_frontier_for_innovation.

[15] M. McLuhan. *Understanding Media: The Extensions of Man*. Cambridge: MIT Press, 1994.

[16] J. J. Meier. *Mathematically significant: A review of awards in mathematics with emphasis on the Fields medal*. 2012. URL: http://www.personal.psu.edu/jjm38/blogs/meier/fieldsmedal.pdf.

[17] American Mathematical Society. *Frank Nelson Cole Prize in Number Theory*. 2012. URL: http://www.ams.org/profession/prizes-awards/ams-prizes/cole-prize-number-theory.

[18] *STAR METRICS*. URL: https://www.starmetrics.nih.gov/Star/Participate.

[19] *Writing by hand better for learning, study shows*. Jan. 11, 2011. URL: http://news.medill.northwestern.edu/chicago/news.aspx?id=177291.