Impacts of the Digital Ocean on Education

Authors
Kristen E. DiCerbo
John T. Behrens

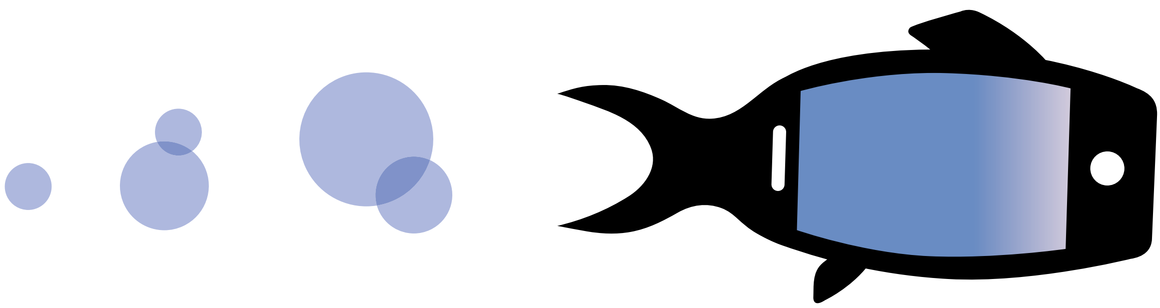
Foreword by
Sir Michael Barber

February 2014

Impacts of the Digital Ocean on Education

Kristen E. DiCerbo & John T. Behrens

February 2014



About the Authors

Kristen E. DiCerbo

Kristen DiCerbo is a Principal Research Scientist at the Center for Digital Data, Analytics, and Adaptive Learning at Pearson. Her research programme focuses on the use of interactive technologies in learning and assessment and the communication and use of data to inform teaching and learning decisions. Kristen investigates how to use evidence from learner activity in games and simulations to understand what learners know and can do. Prior to joining Pearson, Kristen provided research support to the Networking Academies at Cisco and worked as a school psychologist in a local school district in Arizona. She received her Master's Degree and PhD in Educational Psychology from Arizona State University and her BA in Psychology and Sociology from Hamilton College.

John T. Behrens

John Behrens is Vice President of the Center for Digital Data, Analytics, and Adaptive Learning at Pearson. John promotes, researches, and creates digital learning experiences that combine computational and statistical methods with psychological principles to improve learning and education. His methodological interests include the logic of data analysis and the application of the lessons of social science to improve data analysis. Prior to Pearson, John directed learning product development and analytics in the Cisco Networking Academies, serving over 10,000 schools across 165 countries. He obtained his PhD in Psychology in Education at Arizona State University, where he later held the position of tenured Associate Professor. John is also an Adjunct Assistant Research Professor in the Department of Psychology at the University of Notre Dame (Indiana, USA) from where he received a BA in Psychology and Philosophy.

About Pearson

Pearson is the world's leading learning company. Our education business combines 150 years of experience in publishing with the latest learning technology and online support. We serve learners of all ages around the globe, employing 45,000 people in more than 70 countries, helping people to learn whatever, whenever and however they choose. Whether it's designing qualifications in the UK, supporting colleges in the US, training school leaders in the Middle East or helping students in China learn English, we aim to help people make progress in their lives through learning.

Introduction to the Series

The Chief Education Advisor, Sir Michael Barber, on behalf of Pearson, is commissioning a series of independent, open and practical publications containing new ideas and evidence about what works in education. The publications contribute to the global discussion and debate big "unanswered" questions in education by focusing on the following eight themes: Learning Science; Knowledge and Skills; Pedagogy and Educator Effectiveness; Measurement and Assessment; Digital and Adaptive Learning; Institutional Improvement; System Reform and Innovation; and Access for All. We hope the series will be useful for policy makers, educators, and all those interested in learning.

Creative Commons

Sample reference: DiCerbo. K. E. & Behrens, J.T. (2014) *Impacts of the Digital Ocean*, London: Pearson.

Pearson 2014
The contents and opinions expressed in this paper are those of the authors only.

ISBN: 9780992422028

Acknowledgements

We wish to express our thanks to all of those who have helped shape our thinking on the topics discussed here. Our extensive collaborations with Dr Robert Mislevy led to many insights into the structure of inference and its relationship to data. We are also influenced by others working on similar issues, including Dr Valerie Shute and Dr Russell Almond, who have generously shared their time and thoughts. Practical development experience with many colleagues in the Cisco Networking Academy provided a picture of what is possible with forward-looking, technology-enabled teams. We are especially grateful to Dr Richard Halverson at the University of Wisconsin for his feedback and comments on an earlier draft. We would also like to thank Dr Kimberly O'Malley of the Research and Innovation Network at Pearson, who has supported and encouraged this work. Finally, we are grateful to Elpida Ahtaridou for her tireless work and good spirits in getting this work from initiation to final product.

Contents

Foreword by Sir Michael Barber
Preface
Executive Summary i

1 Introduction 1

2 The Digital Revolution 4

3 Data in the Wild 6

4 From Data to Decisions 10

5 Making Decisions 11

6 Activity in the Digital Ocean 14

7 Games in the Digital Ocean 20

8 Potential Sandbars 23

9 Seeing with New Digital Eyes 25

Summary 27
Bibliography 28

Foreword
by Sir Michael Barber

The "digital ocean" that this paper introduces is coming. Just as "big data" is transforming other industries such as insurance, finance, retail, and professional sport, in time, it will transform education. And when it does, it will resolve some long-standing dilemmas for educators and enable that long-term aspiration for evidence-informed policy at every level, from the classroom to the whole system, to be realised.

The dilemmas are well known. Until recently, getting data on outcomes meant formal testing of students. Formal testing, however, is often expensive, time-consuming, and burdensome. It also only provides periodic and, at best, some, but not all, of the outcomes that educators are interested in.

In 1997, when I joined the Department for Education in England, one of the first campaigns I ran within the bureaucracy was for a national data system with individual student level data. Such data would enable us to track students' progress through the system, and, in turn, allow us to identify which schools and districts were succeeding and which were facing challenges. In addition, the data would put us in a position to pinpoint which groups within schools were succeeding and which were poorly served. It took three years to get the database established, and, ironically from my point of view, it began to provide some significant insights just as I was leaving the Department in 2001. Nevertheless, it has had a significant impact in terms of informing policy and enabling the targeting of support and resources.

That was then. Now, the promise of the digital ocean is greater and potentially transformative. Once much of teaching and learning becomes digital, data will be available not just from once-a-year tests, but also from the wide-ranging daily activities of individual students. Teachers will be able to see what students can and cannot do, what they have learned and what they have not, which sequences of teaching have worked well and which haven't - and they will be able to do so in real time. Personalised learning at scale will become possible. And students themselves will be able to get feedback in real time. *"Students taking charge of their own learning"* can stop being simply an oft-repeated phrase and become a daily reality.

Of course, the digital ocean alone will not ensure that these positive developments actually occur. Other evolutions alongside the data revolution will be essential. For example, schools will need to have digital materials of high quality, teachers will have to change how they teach and how they themselves learn, and teachers, principals, and system leaders will have to ask the right questions to get intelligible answers. Otherwise, educators and leaders could drown into the digital ocean rather than swim in it.

DiCerbo and Behrens have done everyone interested in improved educational outcomes a great service, by setting an aspirational vision of what success might look like: they see teaching, learning and assessment as different aspects of one integrated process, complementing each other at all times, in real time; they see sophisticated student profiles allowing teachers and students to make informed and precise decisions about next steps; and they see more complex educational outcomes, such as inter- and intra-personal skills, becoming assessable, teachable, and learnable.

But there is much more to do before this vision becomes a reality across education systems. To begin with, we need further research that brings together learning science and data science to create the new knowledge, processes, and systems this vision requires. Also, we cannot underestimate the technical challenges of validity and reliability, as well as those related to the technology. Nor can we ignore the structural and cultural changes that are required at a systemic and a pedagogical level, which will not come easy, as Michael Fullan and Katelyn Donnelly point out in Alive in the Swamp. Be that as it may, the aspiration to meet these challenges is right, and the "digital ocean" will undoubtedly become a key ingredient of our capacity to improve learner outcomes, not incrementally, as has happened over the past 30 years, but at the dramatic rate that the challenges of the 21st century require.

We hope this paper, and others in our series, will further support discussions on how we can effectively support all learners in the 21st century. The paper does not provide all the answers. It does, however, present a clear vision for the future of teaching, learning, and assessment that can positively impact on learning outcomes.

Preface

The ideas we write about here are the culmination of more than 20 years of our very practical participation in the learning and assessment fields: working in schools, building curricula, developing simulations and games that were distributed to millions of learners, and experiencing the influence of technology on our everyday lives.

We believe the ability to capture data from everyday formal and informal learning activity should fundamentally change how we think about education. Technology now allows us to capture fine-grained data about what individuals do as they interact with their environments, producing an "ocean" of data that, if used correctly, can give us a new view of how learners progress in acquiring knowledge, skills, and attributes.

Our use of the term "digital ocean" to describe the vast amount of data that is available from interactions with digital tools can be traced to one specific assessment-related conference. We were the last speakers on the last day and had heard presentation after presentation discussing how to transition from paper-and-pencil tests to computer-based assessments that looked a lot like those traditional paper-and-pencil tests. It reminded us of the research on technological revolutions which points out that, in early stages of such shifts, new technologies are interpreted from the socio-cultural perspective of the preceding paradigm. This leads to anachronisms such as early automobiles, which placed the driver high above the street – in the same position as the driver of a horse-drawn carriage.

We wondered how it was that our conversations were still so dominated by discussions of individual "items" and "drop in from the sky" assessments unrelated to learning activity. Having been either "born on the boat" or born as a "digital native," we found the juxtaposition of rich daily digital experiences with pre-digital assessment paradigms jarring and confusing. We spent the two hours before our presentation furiously re-writing it to include the big ideas you see here and coining the terms "digital desert" and "digital ocean."

Assessment is ultimately about gathering evidence to support claims we wish to make about learners. We wrote this paper to enumerate the ways in which the digital revolution impacts nearly every element of this process, including the question of what constitutes evidence. Most importantly, the changes we are anticipating and suggesting require new ways of thinking and, in many cases, a new language.

However, the title of this paper is *Impacts of the Digital Ocean on Education*, not *Impacts of the Digital Ocean on Assessment,* because the changes we describe break down the barriers between learning and assessment. This requires not just rethinking assessment, but also rethinking how we view and design teaching and learning activity. As will become apparent, we do not have all the answers - and may not even have all the questions. However, we hope this paper will prompt readers to think differently about learning, inference, and technology in the 21st century.

Kristen DiCerbo
John Behrens
February 2014

Executive Summary

The devices and digital environments with which we interact are designed to record and store experiences, thereby creating a slowly rising ocean of digital data. We can imagine schools and individual learners using this "digital ocean" to inform decisions about learning. As learners learn, they are able to collect information about their activities and get feedback about what they know and can do. Learning can occur in formal and informal contexts, and data can be drawn from both. In the digital ocean, we would expect to see data from all types of activities and contexts used to create persistent learner profiles, which could then be used to recommend future activity.

Our new ability to record, store and process information from learners in numerous non-testing situations suggests a paradigm shift in assessment to include the following:

- focus on a broad range of attributes versus measuring narrowly defined knowledge and skills
- assessment from in vivo naturalistic tasks vs. via pre-made tests
- integration of data over activity and time as opposed to over singular events
- detailed tracking of context outside of testing situations
- dissolution of current distinctions such as "informal" vs. "formal" learning
- collection and permanence of learner profile data to make ongoing, intelligent recommendations.[1]

The data of the digital ocean is not simply more data as we knew it in the pre-digital era. Several aspects of data generation, storage, transfer and use make the new data qualitatively different from that of the past. It is ubiquitous (coming from all manner of activity) and persistent, and it reflects social connection. However, the data itself is only a starting point that is necessary, but not sufficient to transform education - or any activity. Data is only a representation or symbol of what happens in the world.

[1] DiCerbo & Behrens 2012

In most contexts, the goal of data collection and analysis is to provide insight and inform decisions. In classrooms, the fundamental decision of the learning process is what the learner should do next. In the paper, we break this decision into a series of steps, beginning with an activity. Learner interactions with activities generate data that can be analysed for patterns. Such data lead to updates of the learner's profile maintained by the teacher, parent, computer or learner themselves, which in turn informs subsequent decisions or recommendations. Performance in individual activities can often provide immediate feedback (*"you forgot to carry the eight"*) based on local pattern recognition, while performance over several activities can lead to profile updates, which can facilitate inferences about general performance (*"you are good at multiplication"*). Profiles can be used within an activity system (class, computer programme, etc.) or across systems, if persistence and process connection is planned. The shift toward the digital ocean opens up new possibilities for conceptualisation and action in the above process, particularly in the following ways.

Interaction

With technological advances, it is now possible for computers to score a much wider range of responses. This allows for the subsequent expansion of the types of activities that can be used to characterise the knowledge, skills, and attributes of learners and, hence, the forms of assessment we can allow.

Data

Technology provides us with the space to expand our thinking about evidence, which, here, is what we can collect from interactions between learners and learning material. Digital systems allow us to capture stream or trace data from learners' interactions, which will expand our ability to understand learners' processes when solving problems, not just their final products.

Pattern Seeking

Human and computer analysis of these new data types can uncover new patterns that may provide evidence about learning.

Immediate Feedback

Once information is gathered about an activity or 'work product,' and an interpretation is assigned to it, that information can be used to provide immediate feedback to learners and/or their teachers about their performance on that specific activity.

Learner Profiles

Having more data also helps us build better models of learners' knowledge, skills, and attributes. The data helps us "tune" these models, ensuring that the relationships between what we observe from learners and the inferences we draw about their proficiencies are accurate and valid.

Activity Recommendation

Data ultimately enables us to take a learner's profile of knowledge, skills, and attributes and determine the best subsequent activity for that learner's effort to meet a particular goal (for example, learning, motivation, or assessment).

While the digital ocean requires changes to technological systems and to some of our statistical and data analysis procedures, perhaps the biggest change required is a change in thinking. Categories of thinking and acting have often been based on the limits and artefacts of the pre-digital era. As we move forward in the age of the digital ocean, we must step back and consider not only new, specific technologies, but also new relationships to existing categories. Specifically, we will need to move:

- From **items in isolation** to **activity in context.** Rather than thinking of a multitude of individual, isolated items, the digital ocean encourages us to think about integrated activity. From this perspective, rather than posing questions, activities request action.
- From **assessment isolation** to **educational unification.** The costs of collecting, analysing, and making inferences from data decrease with technological advances. This means that more learners and educators should be able to make use of standardised deployments of systems in everyday use. Such systems employ statistical and data-based tools to provide feedback on performance and understanding of themes relevant to classroom teaching and learning. The historic separation between learning and informal inference for classroom concerns can be left behind.
- From **individual paradigm** to **social paradigm.** As technology facilitates movement toward collecting information from more natural interactions, the social nature of everyday lives will become apparent in the data collected. While this will present technical challenges to assessment, it will also allow for more authentic experience and a closer relationship between assessment tasks and life outside the classroom.

Many aspects of the story presented here are aspirational, and some are cautionary. However, many are based on existing experiences of educators and learners around the world, who, as digital natives, are moving forward in their embrace of digital technologies and who, in a few more years, will be unaware of any shift having taken place at all. For those of us who have emerged from the pre-digital era, the challenge is to move beyond the understanding of new technology as a new means to acquire old ends and to re-invent our conceptualisations to take advantage of a digital-first, data-first world.

1 Introduction

We live in an era of remarkable social change brought about by the pervasive use of new technologies. Our laptops, phones, and tablets help us to create our work and play, extend and communicate with our social network through text, voice, and video, and manage the errands of our daily lives. For many of us, network-enabled digital devices are part of the fabric of daily life, and memories of life before the World Wide Web, mobile phones, and smart appliances are either fading away or never existed. While we are aware of the greater efficiency, connectedness, and fun of this emerging digital world, we likely have only fleeting memories of the games we played, the searches we conducted, or the exact words we typed or spoke. By contrast, the devices we interact with are typically designed to record these fleeting experiences, creating a slowly rising ocean of digital data. This emerging digital ocean,[2] when combined with appropriate analysis and standards for use, opens the door to new types of naturalistic observation and inference that could help us to understand and improve ourselves. When extended to education, we expect such changes to advance our learning and our stewardship of the learning of others.

We live in an era of remarkable social change brought about by the pervasive use of new technologies

Emerging from the Digital Desert

We can see the digital ocean of data slowly rising from our post on the edge of a historical era we call the "digital desert." In the digital desert, data collection and storage was expensive, limited, and isolated. School offices needed large filing cabinets just to hold the folders containing learner score summaries from end-of-year summative assessments. Teachers had their own physical records of learner experiences and their own mental models of each student's strengths and weaknesses. However, these records and models were not portable, easy to share, or quickly analysable. Assessment activities could be personalised by the teacher and constructed to be interactive, but the data from classroom activity typically remained local, and, often, ephemeral. While teachers and learners were working together, other individuals, including the next year's teacher, the school principal, and government and/or state education officials had little chance to understand those processes and interactions, and no systematic large-scale way to monitor outcomes. Even a small portfolio of paper-based work products for each learner to be passed to the next year's teacher would require a large amount of storage space and would be difficult to share with anyone else. In the digital desert, individuals worked in a paper world that allowed for teacher control, flexibility, and interpersonal negotiation, but the final transaction was fleeting.

[2] DiCerbo & Behrens 2012

In response to this situation, educational institutions created systems to monitor the outcomes of educational interactions through end-of-year large-scale summative assessment.[3] These assessments, widely in use today, require activity supplemental to the classroom activity so that information can be gathered about learners' knowledge, skills, and attributes. While removed from the details of the class, and at a level of detail often misaligned with teaching and learning decisions, they provide a form of summary that allows information to travel to users who are distant in time and place. They also allow, theoretically, independent individuals to score the data, providing an outside lens with which to view outcomes.

Since these assessments have to collect information about numerous different, broad domains, efficiency became a key goal in large-scale assessment. The computing power of optical scanners and emerging computers provided new economies of scale, but only for tasks with fixed-response formats (such as multiple choice questions). Information from these assessments was reported back to schools and parents via paper reports. The absence of mobile computing devices and information networks in the digital desert inhibited the movement and comparison of data across social situations and groups.

At the Edge of the Digital Ocean

Looking into the future digital ocean, we can imagine schools and individual learners harnessing ubiquitous and naturally generated data to support decisions about learning. In this emerging space, learners use a digital intelligent math tutor that records each step in a learner's response to a question, the scoring of each task, hints requested, and resources used by the learner. Learning is personalised based on learners' knowledge states and trajectories, and the creators of the systems improve them over time as data helps them to understand the processes of learning. During science lessons, learners engage in a variety of physics simulations that allow them to manipulate basic items such as balls, inclined planes, and levers in response to requests to demonstrate their knowledge of concepts such as force and motion. Learners record predictions of what will happen in the simulation in free text. The software records each problem, the placement of each object, the time spent manipulating each item, and the learners' predictions. The learners' writing is automatically scored for evidence of major concepts using mathematical models to find relationships between the meaning of words in text.[4] In an after-school programme, learners collaboratively play a game in which they design various amusement park rides by applying physics concepts, such as the conversion of potential energy to kinetic energy in a roller coaster. Data recorded include: amount of time learners spend on the system, rides designed, time spent on each ride design, and design choices made. This data is accumulated over time to provide learners with estimates of their proficiency and to inform the game itself as it determines which scenarios to unlock.

[3] Piety 2013

[4] Landauer, Foltz & Laham 1998

As with end-of-year large-scale summative assessments, these activities offer standardised presentation and scoring across contexts. However, they are more personalised than summative assessments, and they deliver ongoing interactions that break down the barriers between assessment, teaching, and learning. As learners interact in digital environments, they are able to collect information about their activities to understand what they know and can do. Learning can occur in formal and informal contexts and actionable data can be drawn from both. In the digital ocean, we would expect to see data from all of these types of activity used, along with information about context, to create persistent learner profiles. If a learner plays a collaborative game after school, conducts a simulation in class, and completes homework in an online intelligent tutoring system, all related to the concepts of force and motion, each of those activities should contribute to an updated summary of the learner's proficiencies, independent of time and location.

We are on the edge of great change as data collection is increasingly embedded in modern life

These developments have the potential to inform us about patterns and trajectories for individual learners, groups of learners, and schools. They may also tell us more about the processes and progressions of development in ways that can be generalised outside of school. Our previous studies have shown that our new ability to record, store, and process information from learners in non-testing situations suggests a paradigm shift in assessment to include the following:

- focus on a broad range of attributes versus measuring knowledge and skills
- assessment from in vivo naturalistic tasks vs. via pre-made tests
- integration of data over activity and time as opposed to over singular events
- detailed tracking of context outside of testing situations
- dissolution of current distinctions such as "informal" learning
- collection and permanence of learner profile data to make ongoing intelligent recommendations.[5]

In summary, we believe we are on the edge of a great change, not only because of the increasing amount of available data, but also because the means of data collection are increasingly embedded in the fabric of modern life. This change offers the possibility of personalisation through unobtrusive support to the learner and content recommendation and presentation, while also allowing us to learn more about learning and improve educational processes.

[5] DiCerbo & Behrens 2012

2 The Digital Revolution

For most individuals in modern society, daily activity increasingly involves interaction with digital devices that also act as sensors in larger technology infrastructures. Today, practically any object can be tagged with a sensor that can gather information about the immediate environment and share information over the internet. Sensors are embedded in so many objects that it is estimated that by 2018, there will be 7.6 billion people with 25 billion internet connected devices in the world.[6] Information from these sensors is communicated to people via a wide range of applications, from parking spaces that provide data about when they are empty, to remote operation of household appliances, to health monitoring of chronically ill patients. The most commonplace examples, however, are the mobile "phones" that millions of people carry. Software embedded in the phones is often designed to capture a user's location in the Global Positioning System, which provides data on factors such as speed, choice of routes, and affinity for destinations. The combined data of multiple users can provide the basis for inferences on traffic patterns and congestion. Patterns of mobile phone use also provide information on social and business relationships. The accelerometers built into devices for purposes such as gaming allow for the collection of data about the movement of users.

While smartphones are the most common computing device available to individuals in some locations, in many portions of the educational community learners interact primarily through general computing devices such as laptop and desktop computers. In this context, sensors are embedded into software, which is typically the data collection and management interface for the user. When working with online software through a web browser, much of the operational management may occur remotely on computers that are centrally managed for software updating as well as data collection and analysis. In other words, the local computer shares information about an activity with a remote computer that gathers information from many local computers. This magnifies the scale of data collection, frees owners of local computers from having to update and re-install software, and may lower operation costs.

Within the educational world, certain segments of the learner population are already shifting large portions of their educational activities into interactions with digital systems such as intelligent tutors,[7] learning management systems that support online collaboration, and, most recently, Massively Online Open Courses (MOOCs).[8] These environments are typically designed with digital instrumentation in mind to support learning and personalisation as well as the use of learning analytics[9] to support administrative functions. Among other learner segments, the

[6] Evans 2011
[7] Feng & Heffernan 2006
[8] Daniel 2012
[9] Siemens & Long 2011

use of technology in the classroom is not so widespread, and school data systems tend to be compartmentalised.[10] However, 85% of teachers in the United States (US) report using technology in their classrooms daily.[11] While the adoption of technology in the classroom has been a relatively lengthy process, the most recent statistics available indicate that, in 2010, 97% of US teachers had at least one computer in the classroom every day, and the ratio of learners to computers in the classroom was 5.3:1. Further, 69% of learners used computers sometimes or often during teaching and learning time.[12]

[10] Means, Padilla & Gallagher 2010
[11] Gates Foundation 2012
[12] Gray, Thomas & Lewis 2010

3 Data in the Wild

Most computers and sensors create activity streams that represent activity over time, in the sequenced order in which it occurred. The term "activity stream" comes from social networking – Facebook's newsfeed is a well-known example. However, the concept can be expanded to include the recording of very fine-grained activities, such as specific mouse clicks on a site, and recording of events across applications. The power of these streams lies in their ability to record change as it occurs. For the purposes of education, they have the potential to indicate changes in learning, motivation and other characteristics of interest as they happen. The granularity of the information can help pinpoint the moments when learning occurs or a learner's approach to a problem changes.

The data of the digital ocean is not simply more data as we knew it in the pre-digital era (e.g., surveys, questionnaires, and experiments apart from the field). Several aspects of data generation, storage, transfer, and use make the new data qualitatively different from the data of the past. In this section we describe how ubiquity, persistence, and connectedness of the data, – meaning connecting data of different users and from different contexts, - change the landscape of learning possibilities.

Detail from Ubiquity

Historically, educational data has largely been generated in the context of isolated events, such as a test given under specific conditions separate from teaching and learning. In the digital ocean, data comes from all manner of activities that occur throughout the day both inside and outside of the classroom. The data of the digital ocean is larger and has more impact because it is recorded from the actions of our daily life rather than in isolated testing situations. This proximity allows us to more clearly understand the micro-patterns of teaching and learning by individuals and groups.

To support learning, we need not more of the same data, but different kinds of data

Teaching and learning is a specific social process designed to change behaviour within the learning setting. While its ultimate goal is to transfer the learnt ideas and actions to other settings, seldom does a teacher follow a learner around outside the classroom to assess that transfer. In the emerging digital ocean, however, the mobile applications ('apps') that help us manage our finances, health, and schedules provide insights into our behaviour that are supported by data gathered in real time. For example, a number of applications are available that act as running or health training coaches. The accelerometer and GPS embedded in phones provide users with data regarding their level of activity and movement throughout the day. This can be combined with health data from other sources and shared with personal trainers, physicians and others. It can also be fed back into the health application to obtain automated feedback and coaching that is naturally intertwined with the daily life of the individual.

Learners at the Quest to Learn school in Chicago[13] use similar technology to collect data about their activity, which is displayed in a game-like format to support the self-monitoring and acquisition of the health behaviours typically taught in physical education classes. The software helps individuals collect data about their health and activity ubiquitously throughout the day. Thus, it breaks down the limitations of time and space required for specific class activity and opens the realm of daily life to the goals of the "gym class."

In the current educational context, we seldom introduce information about out-of-class activity into our understanding of learners. However, it is common for learners to report home reading activities to schools and summer reading programmes. With the rise of ebooks, it would be possible to directly capture not only pages read, but how long learners spent per page, which words they looked up, what text was highlighted or bookmarked, and the difficulty of the reading material. Such information can be used to inform educators about group trends and variations and to point out specific areas that may need to be discussed. This information could be combined with detail on in-school reading to produce more accurate and shareable measures of reading skill. It could also be used to inform parents in near-real time and to make suggestions for activities learners can do at home.

New methods of data capture create records of very small interactions with digital systems (clicks, moves, etc.). This highly detailed information allows for the examination of learner processes rather than of final products alone. Examination of learner activity at this more micro-level can reveal sequences and orders of actions, allowing us to compare the effectiveness and efficiency of various paths. Data from these new micro-measurements opens the door to improved theory-building and testing, as illustrated in the improved understanding of learner experiences in digital environments. For example, data has helped researchers gain better understandings of the relationships between boredom, frustration, confusion, and learning outcomes in digital intelligent tutors.[14]

Persistence and Process Connection

While the new devices and applications embedded in our daily lives provide the possibility of using widespread personal and institutional data collection to aid our learning and personal management, it is the storage of this data that opens possibilities for understanding our actions and the actions of groups as well as for improving software to help educational processes.

Consider, for example, the college course placement process in the US. When a learner attends university for the first time, someone needs to determine from the array of different classes which one best matches the learner's level and goals in common areas such as writing and mathematics. Current practice is largely focused on the use of placement exams highly customised to the institution's context. However, these tests treat learners as if they had no history. The advanced learner and the struggling learner may take the same test because the test delivery system knows nothing about them as individuals.

[13] Salen 2012
[14] Baker et al 2010

In such cases, it may be possible to improve placement decisions by bringing in performance data from outside the test, including electronic transcripts, badges or certifications from the completion of levels in a tutoring system, or electronic evidence of course completion. With standards for what should be taught and how proficiency can be measured, we can imagine university placement systems communicating with secondary school teaching systems to transfer detailed records regarding areas of strength and weakness that may not be identifiable in the fast-and-furious placement context. Of course, the school-to-university boundary is not the only context in which education currently loses data. Changing teachers for different content areas or across years in the digital desert can also be a locus of important data loss, especially if the primary forms of data are gross-grained summary measures. While these may be transferable in paper or electronic format, they highlight the fact that more detailed data from class activity, which could be more informative for making teaching-related decisions, is "wrapped up" and left in the previous classroom.

Persistence and connection need to be leveraged across more than just the large borders of educational systems, such as the jump from secondary to tertiary schooling. Storing data over time helps individuals to assess patterns of their own behaviour. For example, a number of educational data analysts have illustrated the power of dynamic visual graphics, employing such tools as Google Chart to convey change over time using moving images.

Persistence and process connection can likewise inform educational theory and teaching and learning. The storage of data over time allows for the accumulation of records that form the basis of educational data mining.[15] This accumulation allows patterns to emerge where, previously, data insufficiencies may have caused failures to reveal interactions between different classes of action and their effects. Behrens has argued that, as the digital ocean increases in size and data-generating tools become more prevalent, the amount of data available will outstrip theories of how to make use of it.[16] This is partly due to the sheer amount that will become available as well as to the fact that new forms of data (such as social network activity, game solving, and rich simulations) are recent phenomena for which there is little detailed theory of action. The new data creates new understandings that call for new analytics in a cause-and-effect cycle.

At present, most of our digital systems are completely separate from each other. For example, games don't talk to digital tutors, and digital tutors don't talk to homework systems. Most aspects of the educational system are unaware of the user's profile of previous experience. Connecting these systems and pulling together information on different types of attributes and at different grain sizes could allow for better understanding of learners. For example, a computer adaptive test (CAT) might be more efficient, if it was able to use a history of previous activity as a starting point in an assessment or tutoring activity.

[15] Romero et al 2010
[16] Behrens 2013

Social Connection

Activities and individuals in the digital ocean are intertwined in such a way that we are no longer looking at the performance of a single person in isolation in a sterile environment but, rather, at interactions between individuals, often in scenarios much closer to those found in "real life," outside the classroom/testing environment. This connectedness, both between individuals and to environments, opens the opportunity for much richer understanding of individuals and learning. Given that learning is a social endeavour situated in particular contexts, being able to capture information that includes these interactions will allow us to obtain better data about learning as it occurs.

For example, in the game Nephrotex, learners play the role of interns in an engineering firm assigned to design a filter for a dialysis machine. Learners work both individually and in small project teams. A faculty member or teaching assistant guides them via in-game email and chat. Researchers have developed methods of analysing the chat logs not only to measure knowledge, skills, values and identity, but also to illuminate the connections between these factors.[17] These very interactions, which are not captured in the digital desert, allow us to make more detailed inferences about learners. In addition, playing the game appears to increase not just learning, but also motivation in groups underrepresented among engineering majors.[18]

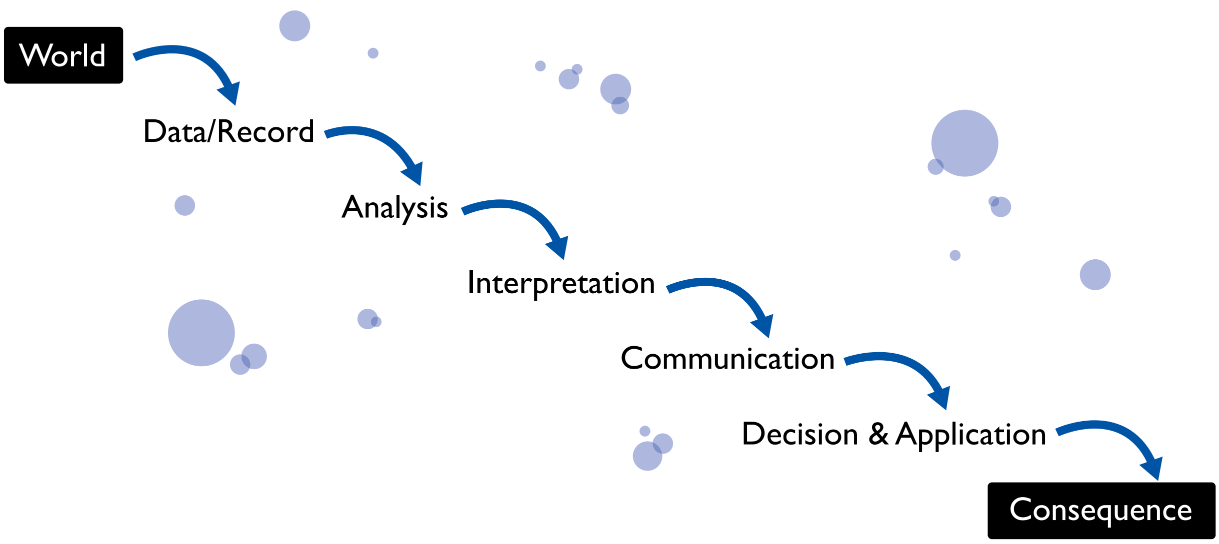
[17] Chesler et al 2012
[18] Chesler et al 2012

4 From Data to Decisions

While the "big data" movement has recently attracted significant attention, as the previous discussion suggests, the focus should not be on just the amount of data gathered. The data itself is only a starting point that is necessary, but not sufficient to transform education. Data is only a representation or symbol of what happens in the world. In most contexts, the goal of data collection and analysis is to provide insight and inform decisions. Accordingly, there is a long chain of reasoning[19] that needs to be considered. Figure 1 provides a contextual overview of the data analytic inference chain. It is important to note that, at every hand-off in the sequence, error can be introduced and possibly amplified later in the process. The causal inferences appropriate to certain forms of data are often more nuanced than they are understood to be at first glance. For example, crime level statistics are frequently reported using both criminal records as well as random surveys of households. This is important because the criminal records created are largely a function of the amount of staffing that occurs in police departments. When more police are added, the number of documented crimes goes up, which may invite the interpretation that more police lead to more crime. The meaning of the records can be properly determined only with the use of background knowledge about the records.[20]

The goal of data science is to inform decisions

Figure 1: Data Interpretation Chain of Inference



[19] Krathwohl 2009
[20] Behrens & Smith 1996

5 Making Decisions

Education experts estimate that teachers probably make around 800 to 1,000 decisions a day.[21] These decisions significantly affect what learners learn; children in the highest reading group may be paced 13 times as fast as those in the lowest, and summative assessment scores reflect this difference.[22] Teaching and learning decisions made on a daily basis about pacing, grouping, and teaching and learning methods have a tremendous impact on learner outcomes. One of the most robust research findings in education is that learner achievement improves dramatically when teachers use information about their learners' learning. In Hattie's meta-analysis of 800 studies, this was the single most powerful (in-school) predictor of learner achievement. He writes: *"When teachers seek, or at least are open to, feedback from learners as to what learners know, what they understand, where they make errors, when they have misconceptions, when they are not engaged – then teaching and learning can be synchronized and powerful. Feedback to teachers helps make learning visible."*[23]

The effect size for this kind of feedback to teachers was 1.13 standard deviations. The data from the digital ocean can help provide this feedback, but the data needs to be tied to decisions and activity.

Learner-Level Decisions

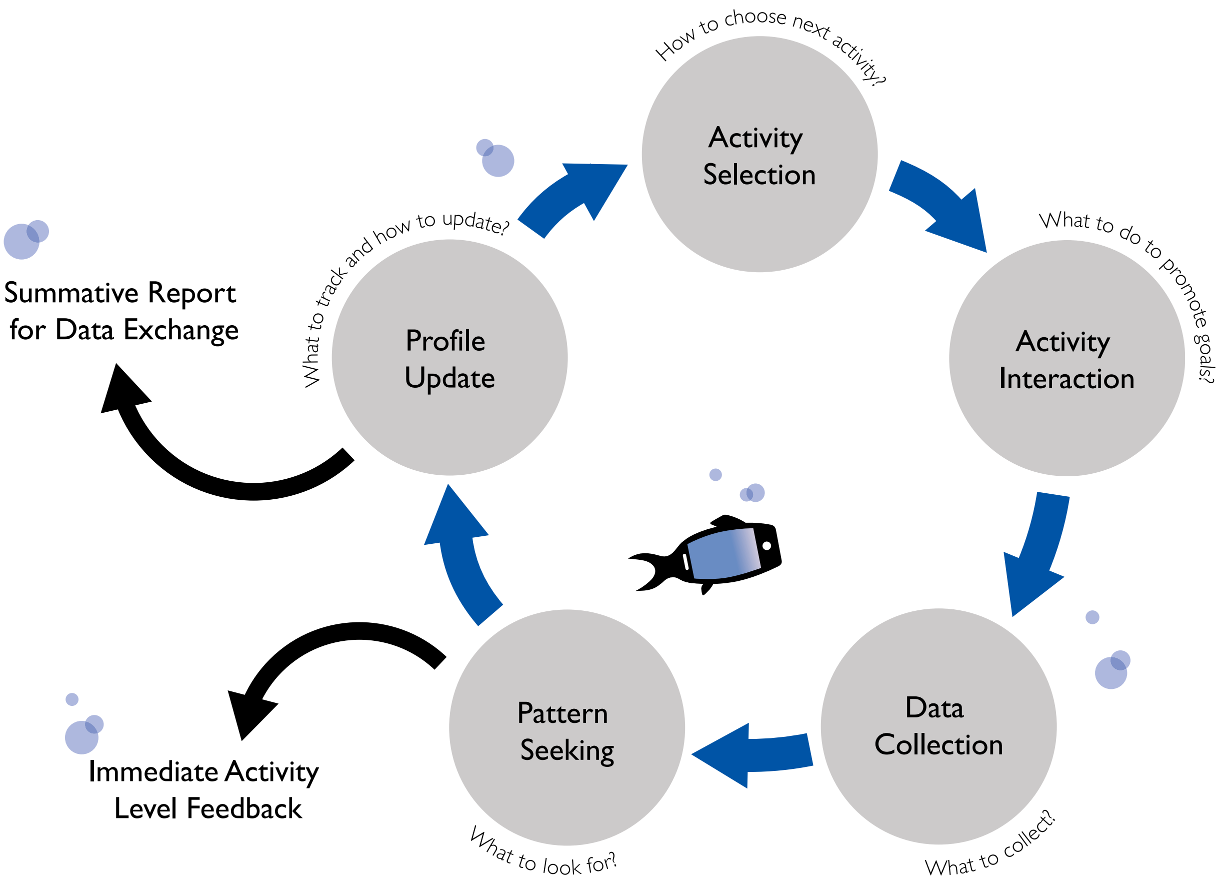
At the most basic level of activity completion, the main decision centres around the following question: What should the learner do next? In order to answer this question, learners and educators engage in a series of further questions:

- What do I (the learner) need to do to complete this activity?
- What are the important outputs of this activity to evaluate?
- What did I (the learner) do well in that activity, and what could I improve?
- How do I (the learner) put the pieces together to get a better picture of my skills over time or across subjects?
- What do I (the learner) do next?

[21] Good & Brophy 2008
[22] Shavelson & Borko 1979
[23] Hattie 2009: 173

In teaching and learning practice, the answers to these questions feed into each other, forming the cycle shown in Figure 2. Interactions generate data that can be analysed for patterns that lead to updates of the learner's profile, which in turn informs subsequent decisions or recommendations. Performance of individual activities can often provide immediate feedback (*"you forgot to carry the eight"*) based on local pattern recognition, while performance over activities can lead to profile updates that include inferences about general performance (*"you are good at multiplication"*). Profiles can be used within an activity system (class, computer programme, etc.) or across systems if persistence and process connection is planned.

Figure 2: The five processes of the general activity cycle



The general activity cycle described above is an extension of the general assessment delivery model[24] from the Evidence Centered Design literature,[25] but it has been applied to a broader array of situations, as described here and elsewhere.[26] Table 1 provides illustrative examples of how a broad array of human and computational activity can be conceptualised from this perspective.[27]

[24] Almond, Steinberg & Mislevy 2002
[25] Mislevy, Steinberg & Almond 2002
[26] Mislevy et al 2012
[27] DiCerbo & Behrens 2012

Table 1: Activity cycle applied to the classroom

	Interactions	Data	Pattern Seeking	Profile Updates	Activity Selection
Teacher in Traditional Classroom	Deliver interactive lecture	Learner verbal responses and non-verbal cues	Are verbal responses correct? Are non-verbal cues consistent with understanding?	Teacher's mental model of learner ability altered	Next activity in lesson plan or topic in curriculum plan adjusted for new understanding
Paper and Pencil Multiple Choice Classroom Test	Fill in bubble with mark	Learner's selected choice	Does selected choice equal correct choice?	Add total number correct and report percent	Complete next question
Computer Adaptive Multiple Choice Test	Read question and choose best perceived answer	Learner's selected choice	Does selected choice equal correct choice?	Add this to a model that estimates learner's ability level taking error and task characteristics into account	Select a question that is most aligned with their estimated ability level
Intelligent Math Tutor System	Enter intermediate steps and final answer to problem using hints as needed	Numeric or text responses along with timestamps and hint use	Does step sequence equal correct sequence? Does final answer match correct answer? How many hints were used?	Add this to a model that estimates learner's ability level taking error and the task characteristics into account	Select a question that is challenging but not frustrating for the learner
Simulation-based Game	Navigate through quests or levels, interact with characters, solve problems	Log file of interactions in the game and final game states	Many possible: sequence of actions to solve problem, time spent on quest, number of events related to problem identification	Add this to a model that estimates learner's ability, interest level or other attributes, taking choices and task characteristics into account	Select a quest that will be maximally motivating for the player

6 Activity in the Digital Ocean

The activity cycle applied to the classroom framework described in Chapter 5 provides a language for identifying aspects of activity traditionally segregated into distinct categories. For example, electronic games and multiple-choice quizzes are typically considered unique and unrelated to each other. While this may be true from a historical perspective of intended purpose, their structures, from the perspective of process and the role of data, have strong parallels. In this section, we review each of these key processes in more detail with respect to how the shift toward the digital ocean opens new possibilities for conceptualisation and action.

Interaction: Types of Interaction Possible

In the past, digital interactions that were used to make inferences about learners' knowledge, skills, and attributes were constrained because automated scoring could only process very limited types of information. The capabilities required to search a complex work product, extract particular features, and apply scoring rules in an automated fashion were beyond the scope of technology at the time, leaving tests as the primary mode of data collection and the form of these tests highly oriented toward fixed responses.

With technological advances, it is now possible for computers to score a much wider range of responses, allowing for an expansion of the types of activities that can be used to characterise the knowledge, skills and attributes of learners and, hence, an expansion of the forms of assessment we can use. We can extract evidence from a variety of work products resulting from a range of activities, including essay writing,[28] computer configuration,[29] and patient diagnosis.[30] We can design activities that require complex responses that are parallel to those that would be required of learners in life beyond the classroom. When we begin activity design, we can begin with a consideration of which "real-world" activities we want learners to be able to perform, rather than what information we can easily score.

Packet Tracer, a computer network simulation tool used in Cisco Networking Academies, specifically links learner interactions with evidence.[31] Teachers and learning designers create activities by setting up initial and final computer networks in the simulation interface which specify the starting configuration and final solution of the activity, respectively. The tool scans the components of the final network (whose modelled features can easily reach into the thousands), and makes a list of all devices and their features. It then presents them to the designers to choose which features and feature values will be compared to the values in the learner's final network solution, after an individual submits their work for scoring. The tool offers additional

[28] Dikli 2006
[29] Rupp et al 2013
[30] Clauser et al 1997
[31] Frezzo, Behrens & Mislevy 2009

facilities for the design of specific end-state connectivity tests, - e.g., whether a test message can travel back and forth across the network,- that are common requirements in "real-world" networking activities. This authoring process results in activities that allow learners to interact with the simulation tool to configure networks just as they would with real equipment.

It is important to recognise a significant shift in the movement from digital desert to digital ocean. The inferential and data collection requirements of large-scale, digital desert testing commonly put testing contexts and practices out of alignment with teaching and learning activity. In this approach, testing is an effort to instrument the learner to provide data, and it requires additional machinery to infer natural behaviour in other contexts from observed behaviour in the sequestered context. However, as the daily activities of life become increasingly instrumented without requiring additional effort, we will move from performances completed primarily to collect data to performances completed for meaningful outcomes with data as a side effect. Increasingly, the activities of studying, practicing, and playing will become sources of data. Tests as we now know them may continue, but in a different role. As we have seen, the digital ocean *"is a world in which data are a side-effect, not the primary goal of interesting and motivating activity, and perhaps a world where 'testing' is a rare event, but assessment is 'in the water.'"*[32]

Data-Tracking of Individual Sequences in Activity Completion

Technology allows us to expand our thinking about evidence, or what we can collect from interactions between learners and learning material. Digital systems allow us to capture stream or trace data from learners' interactions. Figure 3 represents the raw information collected from a traditional multiple-choice test, where the response selected by each individual for each item is recorded. Figure 4 presents a snippet of a typical log file from an interaction with a game, in this case SimCityEDU. This snippet represents a player beginning the game, looking at various information about the state of the city (the tool categories for zones, roads, power, and health), looking at an air pollution map, bulldosing a building, and then checking the air pollution map again.

Figure 3: Typical data captured from a multiple choice test

Student	Item	Option
1001	1	B
1001	2	A
1001	3	C
1001	4	E
1001	5	B
1001	6	C
1001	7	A
1001	8	B
1001	9	E
1001	10	D
1002	1	A
1002	2	A
1002	3	C
1002	4	D
1002	5	B

[32] DiCerbo & Behrens 2012: 302

Figure 4: A typical snippet of a log file from a game

Date Time Stamp	User	Action	Detail
Wed May 15 2013 16:10:06	001_31	GL_Scenario_Accepted	{"name":"Scenario A3 - Large City.txt", "scenarioTime":"00:30"}
Wed May 15 2013 16:10:16	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"zones", "scenarioTime":"00:41"}
Wed May 15 2013 16:10:21	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"roads", "scenarioTime":"00:45"}
Wed May 15 2013 16:10:36	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"power", "scenarioTime":"01:01"}
Wed May 15 2013 16:10:45	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"health", "scenarioTime":"01:10"}
Wed May 15 2013 16:11:20	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"roads", "scenarioTime":"01:44"}
Wed May 15 2013 16:11:26	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"zones", "scenarioTime":"01:51"}
Wed May 15 2013 16:12:12	001_31	GL_Access_Info_Tool	{"name":"Scenario A3 - LargeCity.txt","tool":"jobs","scenarioTime":"02.37"}
Wed May 15 2013 16:12:13	001_31	GL_Access_Info_Tool	{"name":"Scenario A3 - LargeCity.txt","tool":"pollution","scenarioTime":"02.38"}
Wed May 15 2013 16:12:19	001_31	GL_Action_MapLayer	{"action":"opened","name":"Air Pollution Map","scenarioTime":"02:43"}
Wed May 15 2013 16:12:37	001_31	GL_Action_ToolCategory	{"action":"opened","tool":"demolish","scenarioTime":"03:02"}
Wed May 15 2013 16:12:41	001_31	GL_Action_Building	{"action":"viewed","name":"Abandoned Building","scenarioTime":"03:05"}
Wed May 15 2013 16:12:46	001_31	GL_Unit_Bulldoze	{"UGuid":"0xcc9cf003","name":"IndW1D2_Lot96x96_N_0xcc9cf003","Pos": "-801.42,699.60,215.15","scenarioTime":"03:10"}
Wed May 15 2013 16:14:12	001_31	GL_Action_MapLayer	{"action":"opened","name":"Radiation Map","scenarioTime":"04:37"}

Pattern Seeking

Understanding a file of activity such as the one above involves identifying what evidence is important, and then determining how to assign meaning to that evidence. In the traditional multiple-choice assessment, it is clear that the evidence is the option selected. We then assign meaning by comparing it to the predetermined correct response. This correct response is determined by the author of the assessment. In the digital ocean, however, we can explore new patterns in our more complex responses. For example, in the SimCityEDU log, if we were interested in learners' problem-solving processes, we might seek to identify how long players spend gathering information to diagnose the problem in the city. Particular actions could be tagged as "information" activities or "intervention" activities. The data could be pulled out and various measures created, such as number of actions or time spent on information-gathering prior to intervention, or the total proportion of information versus intervention activities. The variability of these measures across players of different skill levels could be examined to determine whether these indicators could be combined into valid measures of problem-solving.

Alternatively, we can ask computers to find patterns. If we have a dataset that we know to be from expert problem-solvers in a domain, and a dataset that we know to be from novices, we can use computer-based statistical analysis to create rules that will best help classify new players into skill categories. When faced with a very large number of potential variables, computers are able to perform pattern identification tasks that are beyond the scope of human abilities. In this way, technology helps us not only to collect information but also detect patterns within it.

Immediate Feedback

Once information is extracted from a work product and an interpretation is assigned to it, that information can be used to provide immediate feedback to learners about their performance on that specific activity. Grading and scoring assignments and exams is a time-consuming and often onerous process. Classroom teachers may spend hour upon hour providing feedback on learners' written work. However, research tells us that feedback is most effective when it is presented as close as possible to the production of the work. Automated scoring allows for immediate feedback to learners, not only on completed work but also on work in progress. For example, the Andes physics tutor indicates whether each step a learner takes is correct or incorrect,[33] the Algebra Cognitive Tutor provides similar correctness feedback and hints if learners get two steps in a row wrong.[34] The Sherlock tutor for electronics troubleshooting provides feedback when learners perform an action that would severely damage real equipment.[35]

Digital environments provide feedback to learners both on completed tasks and work in progress

It should be noted that while immediate feedback is ideal for the learner, it could be invisible to the teacher. In a traditional classroom, the teacher has a direct view of all learner work products and can use those to build mental models of each student's strengths and weaknesses. The use of a digital environment in the classroom puts new pressure on understanding what information teachers need to make decisions. This information must be extracted from digital environments and communicated to teachers in ways that are helpful to the decision-making progress.

Learner Profiles

For many classroom applications, a percentage score indicating the number of correct problems out of a total may be adequate for summarising performance. Such scores are easy to compute and easy to understand. However, percentages also have disadvantages. They summarise how a learner performed only on a particular set of problems. Knowing that a learner answered 35% of questions correct on a test does not actually tell us about their skill, unless we know about the questions on the test. In addition, nearly all assessments contain errors, and resulting unreliability, to a greater or lesser extent. We have more sophisticated statistical models that can help tell us about a learner's levels of proficiency, regardless of the details of the tasks. Two common methods used are called Item Response Theory[36] and Bayesian Networks.[37] Item Response Theory sees the likelihood of a correct response as a combination of both the ability of the person answering and the characteristics of the item. Unlike the percentage example above, Item Response Theory takes the difficulty of an item into account when estimating a person's ability. Bayesian Networks are created by estimating the probability of a person at a given level of skill to perform in a particular way. Then, when a performance from a new

[33] Gertner, Conati & VanLehn 1998
[34] Koedinger & Aleven 2007
[35] Lesgold et al 1988
[36] Van der Linden & Hambleton 1996
[37] Pearl 1988

individual is observed (whether that is a judgment of correctness, a measure of time taken, or a count of events), the Bayesian Network can be used to estimate the individual's likely level of proficiency.

Having more data also helps us build better models; the data allows us to "tune" our models, ensuring that the relationships between what we observe from learners and the inferences we draw about their proficiencies are accurate and valid. As the shift toward the digital ocean provides new ways to estimate relevant attributes, new forms of data also open the door for new opportunities to conceptualise and model attributes. Not only will we make profiles more accurate as we get more data, but having more types of data will also add dimensions to who we consider and model human activity. For example, while current point-in-time summative assessments often focus on core academic constructs (such as mathematics or writing proficiency), data obtained from ongoing class interactions might provide additional information on skills such as time management, perseverance, and teamwork. In the digital desert, these skills are "off the grid" and not likely to be considered part of the flow from data to inference to consequences. In the digital ocean, systems can capture relevant data such as whether a learner is starting homework early or late in the week, and whether learners are studying with massed or distributed practice strategies. The best use of this information may be in the learner's "personal assistant," which would offer study skills and domain proficiency feedback.

Activity Recommendation

The final step that can occur as a result of the new data available is activity recommendation. These recommendations can take the form of adaptation by the computer (offering a new activity) or a recommendation to a learner or instructor to introduce a new activity. Research has suggested that even in schools with developed data systems, the greatest perceived area of need is models of how to connect data to teaching and learning practice.[38] We can use technology to provide these models.

The essence of recommendation is taking the learner profile of knowledge, skills, and attributes, and determining the next activity to best help the learner meet a particular goal. In order to do this, we need to know about learners, available activities, and the goals of the system (learning, assessment, motivation, etc.), as these factors will inform our activity choice. For example, if we are noting flagging engagement, we might choose what we think will be the most motivating activity. If we want to maximise learning, we might choose the activity most within the learner's zone of proximal development. If we want a better estimate of the learner's characteristics, we might choose an activity that will best help us differentiate between novice and intermediate learners.

Some systems are already embedded with such activities. Computer adaptive testing, for example, has been a staple of large-scale assessment for decades. In this approach, the computer tailors the questions to the ability level of the learner. This lowers the time needed for testing (saving expense and enabling faster return to teaching and learning) and avoids presenting the learner with frustrating tasks that may be demoralising.

[38] Means et al 2010

Other Decisions

There are many more decisions that learners and stewards of learning make than simply which activity to do next. Many of these decisions require evidence that ultimately stems from the information above. For example, a classroom teacher might ask, *"How should I group learners?"* Using the learner profiles described above, teachers can specify the aspects on which they would like to base learner groups (for example, two-digit multiplication skill or persistence) and whether they want homogeneous or heterogeneous groupings. A digital system could easily take the stored profile information and suggest groupings based on the teacher's request. Similarly, teachers may want to know, *"How much time should I spend on this unit?"* or, *"What topics do my learners need to review?"* Again, the profile information stored about learners can help them make these determinations.

Similarly, school leaders engaging in curriculum and programme evaluation and making decisions about which programmes to use want to know whether learners are learning as a result of using particular programmes. The ongoing data collection from digital learning environments allows for the monitoring of progress in the short-term, meaning these school leaders do not have to rely for information primarily on end-of-year assessments that have unknown alignment to the programme under review.

7 Games in the Digital Ocean

There has been substantial interest in recent years in games for the classroom, with significant research related to whether learners can learn from games as well or better than they learn from traditional teaching and learning methods.[39] Games also provide a clear example of how data from everyday activities can be used to help us make inferences about learners. Regardless of whether they help learners to learn, games can serve as rich sources of data about them.

1. **Games are already part of the natural activity of learners.** Research conducted in 2008 suggests that 97% of learners aged 12-18 play some kind of digital game.[40] This is a natural activity for them; we don't need to interrupt their lives to collect data.

2. **Games allow us to present "real-world" contexts.** This increases the likelihood that inferences about learners gained from the data will apply outside the classroom.

3. **Games allow us to assess skills we couldn't with a traditional multiple-choice assessment.** Assessment of attributes such as creativity, collaboration, and persistence are possible using data from game-play.

4. **Games provide streams of data that allow us to assess process as well as final product.** We can understand how a learner came to an answer, not just the end result. This allows us to provide better feedback to the learner and gives us a better understanding of how learners progress.

5. **Games share fundamental processes with assessment.** Essentially, this same process holds true for assessments and games.[41] The system presents a problem, the player/learner produces a work product in response, we then evaluate this product based on rules and assign a score. We accumulate these scores into an overall score, and then use that information to present the next task.

[39] e.g. Wouters et al 2013
[40] Lenhart et al 2008
[41] Behrens et al 2006

Game-based Assessment in Action

Shute coined the term "stealth assessment" to describe the process of using information from learners' interactions with digital learning environments to make inferences about their knowledge, skills, and attributes.[42] In contrast to intrusive assessments, which are usually not connected to the learning environment, stealth assessment unobtrusively gathers data from learners' everyday interaction within the teaching and learning environment. Proponents of stealth assessment argue that it allows for the assessment of skills that are difficult to assess with traditional tools. It also provides streams of data that allow for the examination of the problem-solving process rather than simply the end product.

A number of projects have sought to assess science concepts in game-like environments.[43] Shute and her colleagues[44] have designed a game called *Newton's Playground* (NP) to facilitate learning and qualitative assessment of players' physics understanding, persistence, and creativity. NP is a two-dimensional computer-based game that requires the player to guide a green ball to a red balloon. The player uses the mouse to draw simple machines (levers, ramps, pendulums, and springboards) on the screen. The objects "come to life" once the machine is drawn, and the solution is tested to see whether it gets the ball to the balloon. Learners have the opportunity to redraw repeatedly until they are successful. Everything on screen obeys the basic rules of physics regarding gravity and Newton's three laws of motion. The difficulty of a problem depends on the position of the ball and balloon, the obstacles in the way, and the number of machines required to solve it.

The game captures a variety of information about each attempt, including time spent on the level in seconds, number of "restarts" of the level, total number of objects used in a solution attempt, whether the challenge was ultimately solved, and the trajectory of the ball in two-dimensional space. Each of these variables provides useful information about learners' game-play behaviours, which can then be used to make inferences about how well they are doing in the game, their persistence, their creativity, and their current understanding of qualitative physics. For example, if most players use a springboard and ramp to get the ball to the balloon, but a small group uses a lever, the assessment authors might use that as evidence to estimate players' level of creativity. Similarly, the game designers found that an accumulation of the amount of time that players spent on difficult problems in the game correlated to measures of persistence outside the game. In addition, physics knowledge, as estimated by the completion of problems using fewer than three machines, corresponded to physics knowledge as measured by a paper-and-pencil test. Playing the game for four hours across 1.5 weeks led to improved understanding of physics, demonstrating how both learning and assessment took place in the same activity.

[42] Shute 2011
[43] e.g. Halverson & Owen in press
[44] Shute, Ventura & Kim in press; Shute & Ventura 2013

In order to be useful, the estimates of knowledge, skills, and attributes that come from game-play must be good representations of players' knowledge, skills, and attributes outside of games. The long process of establishing validity with evidence from internal structure,[45] response processes,[46] and external measures[47] is underway. Early research suggests that games can be valuable providers of information about a range of player attributes.

Games can be valuable providers of information about a range of player attributes

It should be noted that games are just one type of digital learning environment. There are also intelligent tutors, simulations, and tutorials from which data can be collected. While games are a promising area of research, we should also begin to explore how data from games can be combined with data from these other sources. The accumulation and coordination of information from multiple sources will be what gets us to the digital ocean.

[45] DiCerbo in press
[46] DiCerbo, Frezzo & Deng 2011
[47] Shute et al 2013

8 Potential Sandbars

Even though we believe the digital ocean has the potential for great good, tools are only as good as the processes around them and the alignment of the goals they are intended to achieve. It is important to note four areas of concern that should be considered as the policy discussions unfold.

Privacy and Ownership

In most contexts, individuals have a right to privacy, and in many countries, they have a clear right of access to their data. Laws in this respect have sometimes become outdated, and the general populace accessing educational resources (often minors) may be unaware of either their rights or their responsibilities. Likewise, private and public institutions are often unaware of their legal obligations and may either put privacy at risk, or unwittingly forgo valuable assistance to learners. Organisations need to address these issues proactively by researching and establishing policies in line with governmental regulations and societal expectations. Researchers and other data-seekers must also recognise the limits of data sharing as set by the stewards of that data whether public or private.

Security

While privacy concerns one's *right* to keep information hidden, security may be an even greater concern, as it addresses one's *ability* to keep information hidden. The evolution of computing security has been consistently matched by the evolution of computing security evasions. Most recently, the widespread use of mobile devices opens the opportunity for additional unsecure movement of data and access that could lead to large-scale theft or exposure. Recent publication of large numbers of classified documents related to military and security monitoring in the US highlights the fact that single individuals can have highly disruptive effects, even in systems focused on security issues. In lax environments, the potential for disruption and failure to protect privacy is extreme.

Building Uninviting Spaces

The previous two concerns address technology access and data distribution policies. We also want to emphasise the need to avoid the coerciveness of "the hidden assistant," who chooses all activities for us based on "activity forced for data collection." In short, we should be mindful that instead of reducing reliance on assessments by using data from natural activity, we could inadvertently have the opposite effect and make every activity an onerous exam. To avoid this, we recommend that educators consider a broad ecosystem of activity and data needs, stress the empowerment of the individual through effective use of data, and ensure that the role of affective variables (e.g., joy, agency, self-empowerment, and relationship) is retained in educational activities. These are values nurtured by strong teachers in the early grades, but sometimes forgotten when we seek to optimise specific experiences for specific goals.

Confusing Data and Knowledge

The beauty of data and the excitement of new findings can be alluring and invigorating. This excitement, together with the widespread availability of new analytic tools at little or no cost, opens the door to a tidal wave of "results" and "findings" that may have little or no interpretability. More dangerous, of course, is the possibility of the emergence of numerous impactful, but improper conclusions. For example, as the availability of government-funded educational test score data has increased in the past few years, we have seen an increase in the use of such data. The Apps4VA project promotes the development of tools, which use education data made available from the state of Virginia and represents a positive example of this movement. Unfortunately, in other contexts, we have seen errors in the use and interpretation of such data. One such error is the comparison of learner group performances from year to year using percent-correct metrics. These metrics are typically un-interpretable in that format, as the content of the tests shifts from year to year and the relative difficulty of tests may shift on this scale. For example, some large scale testing programs provide "vertically scaled scores" designed to allow comparison of test scores across grades, which are the types of scores that should be used for these analyses. The technical details available from the qualifying authority should always be consulted.

A similar difficulty is sometimes encountered in which conclusions are justified in their basic form, but are then extended to other, unjustified conclusions. This is the common mistake of confusing correlation with causation, but it is more easily propagated as the sea of data rises and everyone can "become their own captain." For example, analysis of a variety of learning management systems has consistently shown correlations between logging into class on the first day and successful course completion. While this is an empirical fact, we must be mindful that we cannot infer that if we could get learners to log in on the first day, they would be successful. Rather, it is likely that learners who are more motivated are more likely to log in on the first day and do better in the course. Logging in on the first day is an indicator, but not a cause. To move from correlational to causal interpretation would require subsequent analysis that controls "all relevant variables," and, preferably, shows experimental results.

In summary, we believe the benefits for the use of data from the digital ocean can far outweigh the costs. Indeed, the vision we propose is to take best advantage of natural digital activity, not to force more of it. For all the (appropriate) talk about the learning sciences and the data sciences, we believe there should be a corresponding place for the "learning humanities" and the "data humanities," where discourse on ethics, law, human nature, the logic of inference, and ways of understanding the digital age can be brought to bear as well.

9 Seeing with New Digital Eyes

While the digital ocean requires changes to technological systems and to some of our statistical and data analysis procedures, perhaps the biggest change required is a change in thinking. Our categories of thinking and acting have often been based on the limits and artefacts of the digital desert. As we move forward in the age of the digital ocean, we must step back and consider not only new, specific technologies, but also new relationships to existing categories.

> **The biggest change required, if we are to take advantage of the digital shift, is a change is thinking**

Items in Isolation to Activity in Context

Traditionally, we have created assessments made up of "items" or discrete elements that gather information, usually by posing a question, and often by offering a selection of responses. Rather than thinking of a multitude of individual, isolated items, the digital ocean encourages us to think about integrated activity. Rather than posing questions, activities request action. They allow us to extract evidence of attributes rather than just make judgments about correctness. Perhaps most importantly, activities can provide multidimensional information rather than information on a single skill. This way the potential to better assess a wide range of constructs is created, with situations more aligned to the constructs we wish to study. We ultimately want our assessments to tell us how learners will perform in the "real world," where we engage in activities rather than answer a series of discrete questions.

Assessment Isolation to Educational Unification

Our experiences in the field of education suggest a general bifurcation between assessment and the goals, machinery, and expertise behind learning and curriculum. Too often the goals and formats lead not only to different assessment constructs, but also to different skill conceptualisations and the sequestering of different communities of practice. As we have described, current technologies allow for the standardised deployment of systems in everyday use that employ statistical and data-based tools. These tools provide feedback on performance and understanding of themes relevant to classroom teaching and learning. As the costs of collecting, analysing, and making inferences from data decrease with technological advances, more learners and educators should be able to make use of such systems. The historic separation between learning and informal inference in the classroom can be left behind.

Individual Paradigm to Social Paradigm

Learning is collaborative. What we have long known from our emotional experience, we now see in the data of our daily interactions: emails, media posts, tweets, and collaborative workspaces such as wikis. Collaboration in the digital world is helping to solve difficult problems. For example, Foldit is an effort to solve protein structure problems through game-play.[48] Players attempt to discover how real proteins fold, and, by building on the results of others and competing to get the best optimisation scores, they have uncovered the structures of actual proteins that have eluded scientists (and computers). However, very few assessments allow for collaboration; the typical test situation consists of an examinee seated at a desk and being told to *"keep your eyes on your own paper."* As technology facilitates movement toward collecting information from more natural interactions, the social nature of our everyday lives will become apparent in the data collected. While this will present technical challenges to assessment, it will also allow for a more authentic experience and a closer relationship between assessment tasks and life outside the classroom.

[48] Cooper et al 2010

Summary

We are in the midst of a great social shift. Our ability to collect, maintain, and use records of experiences is increasing dramatically each day and for many purposes. When specific goals of data collection present themselves, new technologies are on hand to allow for the collection of complex and ephemeral actions that were difficult, if not impossible to track just a short time ago. This will bring about a revolution in which assessment, teaching, and learning will increasingly make use of authentic and engaging tasks based on simulated environments, social connection, and remote knowledge stores. But, this would only be a technology-enhanced approach to existing constructs. We believe the more compelling story is that the collection of data and the corresponding gains of self-awareness, self-reflection, organisational evolution, and institutional insight are emerging from the bases of our everyday natural activity. Insights are developing from the digital ocean of data that we create in search of rich digital experiences that fulfil our needs for knowledge, fun, and social interaction. It is this fundamental shift from data as the goal of our activity to data as a side effect of our activity that opens new doors to understanding and improving education, and offers the promise of and support for new models of interaction.

Many aspects of the story presented here are aspirational, and some are cautionary. However, they are based on the experiences of educators and learners around the world, who, as digital natives, are moving forward in their embrace of digital technologies, and, who, in a few more years, will be unaware of any shift having taken place at all. For those of us who have emerged from the digital desert, the challenge is to move beyond the understanding of new technology as means to acquire previous ends, and to reinvent our conceptualisations to take advantage of a digital-first, data-first world.

Bibliography

Almond, R., Steinberg, L., & Mislevy, R. (2002). Enhancing the design and delivery of assessment systems: A four-process architecture. *The Journal of Technology, Learning and Assessment, 1* (5). Retrieved from http://ejournals.bc.edu/ojs/index.php/jtla/article/download/1671/1509.

Baker, R. S. J.d., D'Mello, S. K., Rodrigo, M. M.T., & Graesser, A. C. (2010). Better to be frustrated than bored: The incidence, persistence, and impact of learners' cognitive–affective states during interactions with three different computer-based learning environments. *International Journal of Human-Computer Studies, 68* (4), pp. 223–241.

Behrens, J.T. & DiCerbo, K. E. (2014). Harnessing the currents of the digital ocean. In J. Larussen & B. White (Eds.) *Learning Analytics: From Research to Practice*. New York: Springer.

Behrens, J.T. (1997). Principles and procedures of exploratory data analysis. *Psychological Methods, 2*, pp. 131-160.

Behrens. J.T., DiCerbo, K. E., Yel, N. & Levy, R. (2012). Exploratory data analysis. In I. B. Weiner, J. A. Schinka, & W. F. Velicer (Eds.), *Handbook of psychology: Research methods in psychology, 2nd Ed* (pp. 34-70). New York, NY: Wiley.

Behrens, J.T., Frezzo, D., Mislevy, R., Kroopnick, M., & Wise, D. (2006). Structural, functional and semiotic symmetries in simulation-based games and assessments. In E. Baker, J. Dickieson, W. Wulfeck, & H. F. O'Neil (Eds.), *Assessment of problem solving using simulations*, 59–80. New York, NY: Routledge.

Behrens, J.T., Mislevy, R. J., DiCerbo, K. E., & Levy, R. (2012). Evidence centered design for learning and assessment in the digital world. In M. Mayrath, J. Clarke-Midura, D. H. Robinson, & G. Schraw (Eds.), *Technology-based assessments for 21st Century skills: Theoretical and practical implications from modern research* (pp. 13-54). Charlotte, NC: Information Age Publishing.

Behrens, J.T., & Smith, M. L. (1996). Data and data analysis. In R.C. Berliner & D.C. Calfee (Eds.), *Handbook of educational psychology* (pp. 949-989). New York, NY: Macmillan.

Chesler, N. C., Arastoopour, G., D'Angelo, C. M., Bagley, E. A., & Shaffer, D. W. (2012). Design of a professional practice simulator for educating and motivating first-year engineering learners. *Advances in Engineering Education, 3*. Retrieved from http://edgaps.org/gaps/wp-content/uploads/aee-vol03-issue03-01.pdf.

Clauser, B. E., Margolis, M. J., Clyman, S. G., & Ross, L. P. (1997). Development of automated scoring algorithms for complex performance assessments: A comparison of two approaches. *Journal of Educational Measurement, 34* (2), pp. 141–161.

Cooper, S., Khatib, F., Treuille, A., Barbero, J., Lee, J., Beenen, M., Leaver-Fay, A., Baker, D., Popvic, Z. & Foldit Players (2010). Predicting protein structures with a multiplayer online game. *Nature, 466* (7307), pp. 756–760.

Daniel, J. (2012). Making sense of MOOCs: Musings in a maze of myth, paradox and possibility. *Journal of Interactive Media in Education, 3*. Retrieved from www-jime.open.ac.uk/jime/article/viewArticle/2012-18/html.

DiCerbo, K. E. (2014). Game-based assessment of persistence. *Journal of Educational Technology and Society, 17*(1), 17-28. Retrieved from: http://www.ifets.info/journals/17_1/3.pdf

DiCerbo, K. E., & Behrens, J. (2012). Implications of the digital ocean on current and future assessment. In R. Lissitz & H. Jiao (Eds.), *Computers and their impact on state assessment: Recent history and predictions for the future*, Charlotte, NC: Information Age.

DiCerbo, K. E., Frezzo, D. C., & Deng, T. (2011). Substantive validity of a simulation-based game. *Research and Practice in Technology Enhanced Learning, 6*, pp. 161–185.

Dikli, S. (2006). An overview of automated scoring of essays. *The Journal of Technology, Learning and Assessment, 5* (1). Retrieved from https://ejournals.bc.edu/ojs/index.php/jtla/article/view/1640.

Evans, D. (2011). *The internet of things: How the next evolution of the internet is changing everything*. San Jose, CA: Cisco. Retrieved from www.cisco.com/web/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.

Feng, M. & Heffernen, N. T. (2006). Informing teachers lives about learner learning: Reporting in the assessment system. *Technology, Instruction, Cognition, and Learning, 3*, pp. 1-14.

Frezzo, D. C., Behrens, J. T., & Mislevy, R. J. (2009). Activity theory and assessment theory in the design and understanding of the Packet Tracer ecosystem. *International Journal of Learning and Media, 1*(2), Retrieved from www.mitpressjournals.org/doi/abs/10.1162/ijlm.2009.0015.

Gates Foundation (2012). *Technology & effective teaching in the U.S.* Bill & Melinda Gates Foundation. Retrieved from https://edsurge.s3.amazonaws.com/public/BMGF_Innovation_In_Education.pdf.

Gertner, A. S., Conati, C., & VanLehn, K. (1998). Procedural help in Andes: Generating hints using a Bayesian network learner model. *Proceedings of the 15th National Conference on Artificial Intelligence* (pp. 106–11). Cambridge, MA: The MIT Press.

Good, T. L., & Brophy, J. (2008). *Looking in classrooms* (10th Ed.) New York: Pearson.

Gray, L., Thomas, N., & Lewis, L. (2010). Teachers' use of educational technology in US public schools: 2009. First Look. NCES 2010-040. Washington, DC: National Center for Education Statistics. Retrieved from www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED509514.

Halverson, R. & Owen, V. E. (unpublished manuscript). Game-based assessment: An integrated model for capturing evidence of Learning in play.

Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Taylor & Francis.

Koedinger, K. R., & Aleven, V. (2007). Exploring the assistance dilemma in experiments with cognitive tutors. *Educational Psychology Review, 19* (3), pp. 239–264.

Krathwohl, D. R. (2009). *Methods of educational and social science research: The logic of methods*. Long Grove, IL: Waveland Press.

Landauer, T., Foltz, P., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse Processes, 25*, pp. 259-284.

Lenhart, A., Kahne, J., Middaugh, E., Macgill, A. R., Evans, C., & Vitak, J. (2008). *Teens, video games and civics: Teens' gaming experiences are diverse and include significant social interaction and civic engagement*. Washington, DC: Pew Research Center.

Lesgold, A., Lajoie, S., Bunzo, M., & Eggan, G. (1988). *SHERLOCK: A coached practice environment for an electronics troubleshooting job*. DTIC Document. Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA201748.

Means, B., Padilla, C., & Gallagher, L. (2010). Use of education data at the local level: From accountability to Instructional Improvement. Washington, DC: US Department of Education. Retrieved from www.eric.ed.gov/ERICWebPortal/recordDetail?accno=ED511656.

Mislevy, R. J., Behrens, J. T., DiCerbo, K. E., Levy, R., Rupp, A. A., Nugent, R., & Nelson, B. (2012). Design and discovery in educational assessment: Evidence centered design, psychometrics, and data mining. *Journal of Educational Data Mining, 4* (1), pp. 11–48.

Mislevy, R. J., Steinberg, L. S., & Almond, R. G. (2002). Design and analysis in task-based language assessment. *Language testing, 19* (4), pp. 477–496.

Pearl, J. (1988). *Probabilistic reasoning in intelligent systems: Networks of plausible inference*. Burlington, MA: Morgan Kaufmann Pub.

Piety, P. J. (2013). *Assessing the educational data movement*. New York, NY: Teachers College Press.

Romero, C., Ventura, S., Pechenizkiy, M., & Baker, R. S. J. d. (Eds.) (2010). *Handbook of educational data mining*, Boca Raton, FL: CRC Press.

Rupp, A. A., Levy, R., DiCerbo, K. E., Sweet, S., Crawford, A., Calico, T., Benson, M., Fay, D., Kunze, K. L., Mislevy, R. J., & Behrens, J. T. (2013). Putting ECD into practice: The interplay of theory and data in evidence models within a digital learning environment. *Journal of Educational Data Mining, 4* (1). Retrieved from http://www.educationaldatamining.org/JEDM/index.php/JEDM/article/view/23.

Salen, K. (2012). *Seminar*. Presented at the Educational Testing Service, Princeton, NJ.

Shavelson, R. J., & Borko, H. (1979). Research on teachers' decisions in planning instruction. *Educational Horizons, 57* (4), pp. 183–89.

Shute, V. J. (2011). Stealth assessment in computer-based games to support learning. In S. Tobias & J. D. Fletcher (Eds.) *Computer games and instruction.* Charlotte, NC: Information Age Publishers.

Shute, V. J., Ventura, M., & Kim, Y. J. (2013) Assessment and learning of qualitative physics in Newton's Playground. *Journal of Educational Research, 106*, pp. 423-430.

Shute, V. J., & Ventura, M. (2013). *Measuring and supporting learning in games: Stealth assessment*. Cambridge, MA: MIT Press. Retrieved from http://myweb.fsu.edu/vshute/pdf/white.pdf.

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *Educause Review, 46* (5), pp. 30–32.

Van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*. New York, NY: Springer.

Wouters, P., van Nimwegen, C., van Oostendorp, H., & van der Spek, E. D. (2013). A meta-analysis of the cognitive and motivational effects of serious games. *Journal of Educational Psychology, 105* (2), pp. 249–265.

“we believe we are on the edge of a great change, not only because of the increasing amount of available data, but also because the means of data collection are increasingly embedded in the fabric of modern life”