Harnessing the Currents of the Digital Ocean.

Paper Presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA. April, 2013

John T. Behrens

John.Behrens@Pearson.com

April 2013

About Pearson
Everything we do at Pearson grows out of a clear mission: to help people make progress in their lives through personalized and connected learning solutions that are accessible, affordable, and that achieve results, focusing on college-and-career readiness, digital learning, educator effectiveness, and research for innovation and efficacy. Through our experience and expertise, investment in pioneering technologies, and promotion of collaboration throughout the education landscape, we continue to set the standard for leadership in education. For more information about Pearson, visit http://www.pearson.com/.

About Pearson's Research & Innovation Network
Our network **mission** is to spark innovation and create, connect, and communicate research and development that drives more effective learning. Our **vision** is students and educators learning in new ways so they can progress faster in a digital world. Pearson's research papers share our experts' perspectives with educators, researchers, policy makers and other stakeholders. Pearson's research publications may be obtained at: http://researchnetwork.pearson.com/.

Abstract

The digital revolution concerns the shift in human history that allows the transformation of experiential inputs and work products into digital form that can be immediately collected, transformed, moved, stored and computed upon. This shift has already had remarkable social consequences and raises fundamental questions regarding the nature of science and knowledge. In the context of educational research, it raises key questions about the nature of our relationship with data in scientific endeavor and the role of computing systems and computational skills of researchers.

This paper extends the discussion begun by DiCerbo & Behrens (2012) in which they outlined how the societal shift related to the digital revolution can be understood in terms of a shift from a pre-digital "digital desert" to a post-digital "digital ocean". Using the framework of Evidence Centered Design (Mislevy, Steinberg, & Almond, 2002) they suggest that the core processes of assessment delivery can be re-thought in terms of new capabilities from computing devices and large amounts of data and that many of our original categories of educational activity represent views limited by their origination in the digital desert. After reviewing the core ideas of digital desert to digital ocean shift, implications for understanding educational research is addressed in terms of methodological implications of this shift including the role of data in hypothesis generation, the role of data in theory testing, impact of the data to explanation ratio when data size increases dramatically, the impact of continuous data generation and analysis, and the changing role of statistical and computational tools. Implications for graduate training are addressed throughout. The paper concludes with a note of caution.

Keywords: data, methodology, digital ocean

Harnessing the Currents of the Digital Ocean.

Recently, DiCerbo and Behrens (2012) suggested the term “digital ocean” to describe the emerging reality of ubiquitous and unobtrusive data generated from the use of digital devices in daily life in contrast to the pre-digital world of expensive and relatively rare data which they characterize as the “digital desert”. While originally formulated in the context of the impact of these shifts on assessment argument and use, we extend the discussion to the broader context of data-based research in general. This is accomplished in five sections each of which touch on a shift in perspective or activity that is part of the change as we understand it. In the first section the experiential aspects of the shift are discussed following DiCerbo and Behrens (2012). Next, conceptual shifts in understanding educational assessment and educational research data are suggested to provide appropriate conceptual tools for the new and emerging realities. The third section discusses shifts in generation and storage of data. A fourth section discusses issues related to the organization and conduct of research given these shifts and addresses implications for training of educational researchers. Sections relating cautions and conclusions end the paper.

Following the analogy of DiCerbo & Behrens (2012) we are currently on the digital shore: a place in the history of human cultural evolution between the digital desert of the past and the digital ocean of the future. From this epistemic position, discussion of the near past may seem a simple caricature and discussion of the future mere fantasy. However, because the revolution in computing that we are embedded in concerns the transformation of information from physical form and activity to a liquid digital form that can be moved, transformed, synthesized, and acted upon by automated systems (Mislevy, Behrens, DiCerbo & Levy, 2012), it is also a revolution in the nature of human intellectual and cultural history. It will be, we believe, a fundamental lens through which activity will be understood in the next hundred years, in the same way questioning the role of the individual and the value of systematic inquiry was a central lens in the age of the Enlightenment.

I. EXPERIENTIAL SHIFT – SENSORS

The starting point for the conversation regarding the shift from digital desert to digital ocean is that, for most individuals in modern society, daily activity increasingly involves interaction with digital devices which, by their nature, also act as sensors in larger technology infrastructures. Massively multi-functional mobile computing devices (often anachronistically also called “phones”) allow the unobtrusive (and sometimes unrevealed) collection and communication of data to numerous electronic aggregation points. Software embedded in the phone is often designed to capture your location in the Global Positioning system from which speed, choice of routes, and affinity for destinations can be learned. Patterns of cell phone use provide information related to social and business relationships. Accelerometers on these devices enable them to be used as game consoles and collectors of other data. An emerging practice of personal data collection is referred to as the quantified self movement (Wolf, Carmichael, & Kelly, 2010; Wolf, 2002). In the area of medical quantified self, the early identification of a heart attack by remote examination of unobtrusive ekg data can allow for pre-critical treatment (Kappiarukudil & Ramesh, 2010) . Children at the Institute of Play (Salen, 2012) use digital collection techniques to track and mange their own activity and health.

While smart phones are the most common computing device available to individuals in some countries, in many portions of the educational community, students interact primarily through general computing devices such as laptop and desktop computers. In this context, the software being used is the basis of the sensor as they are typically the data collection and management interface for the user. In such environments, the product of the interaction is often captured and stored (e.g., the document created or the outcome of the game) as well as the possibility of ongoing process data such as game logs. When working with online software through a web-browser, the bulk of non-display computing can occur on remote computers that are centrally managed for software updating as well as data collection and analysis. This intensifies the scale of data collection possible.

Within the educational world, certain student segments are already shifting large portions of their educational activities into interactions with digital systems such as tutoring systems (Feng & Heffernan, 2006), learning management systems that support online collaboration, and most recently, Massively Online Open Courses (MOOCs; Daniel, 2012). These environments are typically designed with digital instrumentation in mind in order support learning and personalization as well as the use of learning analytics (Siemens & Long, 2011) to support administrative functions as well.

These technological shifts in sensing, however, would be of little concern if it were not for concomitant shifts in levels of use of digital devices by the general public and the dramatic movement in the use of digital devices for a broad range of daily activity including social communication, entertainment, play activity, broad ranges of commerce, as well as learning for broadly educational purpose and focused search and retrieve activities. One implication of these shifting patterns of activity discussed by DiCerbo and Behrens (2012) is that digital learning activity and thereby digital learning data are able to occur with relatively few constraints of time and location. The student who wants to learn typing or another skill during their “after-school” time has the opportunity to access a broad range of open educational resources (OERs) that may or may not collect or transmit data. Likewise, the use of many “informal” online activities is suggested to have positive learning outcomes (Gee, 2003). While it was always well known that students read and learn outside the classroom and that there are positive educational aspects of many “informal” activities (e.g., team sports), the recordability and subsequent research on these genres of activity suggest a unification of understanding activity and a breaking down of pre-digital boundaries between activity clusters. For example, while the concept of homework has always been fluid (e.g., sometimes it can be done in school), the fact that it can be done at any time in any place using network connected computers raises the question of whether that distinction still has much value. Likewise, a student playing an educational game (or a game with educational impact) might obtain proficiency in curricular objectives (thereby relating to the activity as a curricular object), generate and respond to assessment data (relating to it as an assessment object), and have fun and communicate to friends about performance and strategies (relating to it as a social or play object). Accordingly, DiCerbo and Behrens (2012) argue the rise of digital devices and ubiquitous activity raises into question the conceptual boundaries which arose during the pre-digital era of the digital desert.

II. CONCEPTUAL SHIFT – TESTING TO INTERACTIONS

Working in the context of understanding current shifts in understanding educational assessment practices, DiCerbo and Behrens (2012) apply the language of student-system interaction from Evidence Centered Design (ECD; Mislevy, Steinberg, & Almond, 2002) to understand past and current large scale testing approaches. The delivery process described in this literature is articulated in terms of a four-process delivery model (Almond, Steinberg, & Mislevy, 2002). While this model was originally intended to explicate assessment and tutoring system activity, subsequent analyses brought application to games (Behrens, Frezzo, Mislevy, Kroopnick, & Wise, 2006; Shute, 2011). This model suggests four core processes:

- Activity Selection: What activity is to be presented next to the learner/examinee? This process can be based on electronic student profiles or can be based on teacher's human judgment, or other methods.
- Activity Presentation/Interaction: The process of interacting with learner/examinee and obtaining data. The process could include answering a question or completing a complex simulation on a test, completing a level of a game, or completing a practice activity in the course of instruction. Regardless, the result is a work product that can take many forms including the answer to a question, the log of game activity, or the essay written in a project.
- Evidence Identification or Response Processing: The process of identifying observable features of the work product that can be passed to subsequent summary processes. This could include the application of

Latent Semantic Analysis (Landauer, Foltz, & Laham, 1998) or other Natural Language Processing techniques to an essay that results a list of variables with specific values. In the context of multiple choice testing this often means the generation of a specific bit indicating correctness/incorrectness of the response. In such a context it may also be called item-level scoring.

- Evidence Accumulation or Evidence Synthesis: This is the process of summarizing previous smaller pieces of task level information to create a profile of learner states. This could be as simple as adding up all the points assigned to questions on a test to differential weighting of values based on complex statistical models such as IRT (van der Linden & Hambleton, 1996) or Bayesian Inference Networks (Almond, DiBello, Moulder, & Zapata-Rivera, 2007; Pearl, 1988).

A schematic characterization of the four process model is provided in Figure 1.

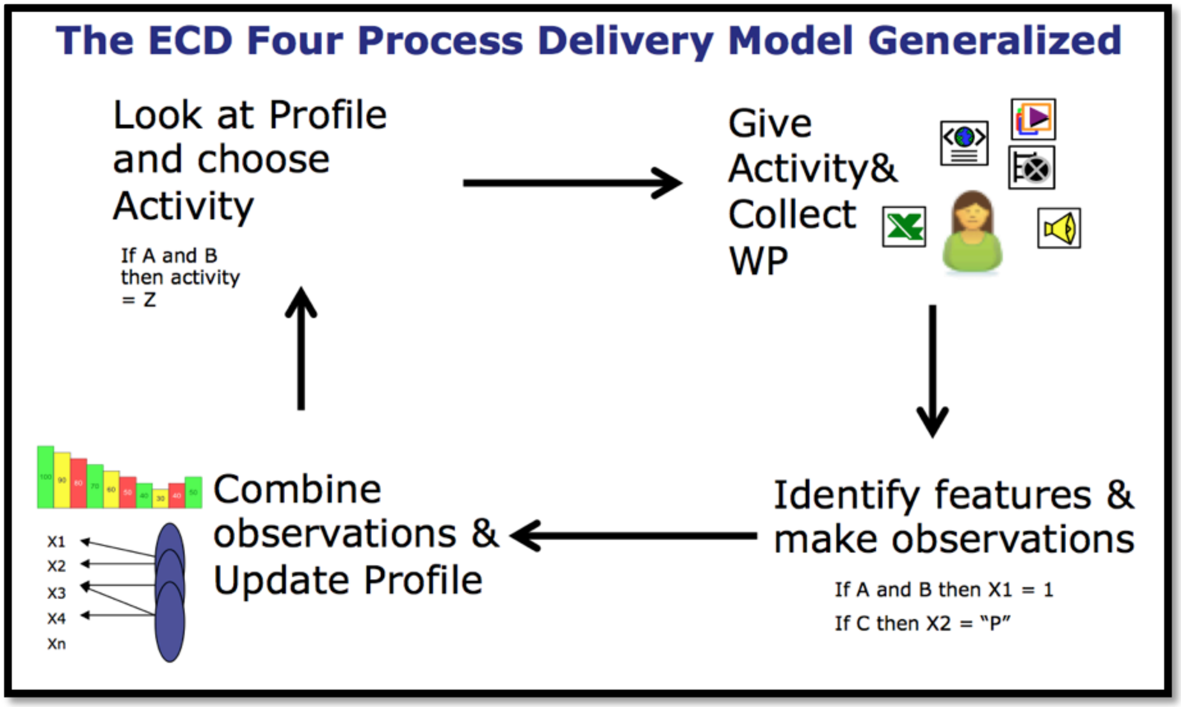

Figure 1. Generalized chacterization of the ECD Four Process model following Almond et al, 2002

DiCerbo & Behrens (2012) point out that while this is a very generalized model (see also Mislevy et al., 2012) that allows for a broad range of activity, the predominant assessment paradigm of the 20th century was as follows:

- Activity Selection: Predetermined ordering of activities in "fixed form"
- Presentation: Questions eliciting fixed responses
- Evidence Identification: Matching of fixed response against fixed answer
- Evidence Synthesis: Add up "correct" responses or differentially weight them using pre-calibrated statistical models

Let's contrast this characterization against a similar analysis of game construction (Behrens et al., 2006):

- Activity Selection: Choose next activity or level based on state of student model
- Presentation: May be simple or complex, possibly providing complex emulation of real or imaginary worlds
- Evidence Identification: May be simple or complex, possibly considering strategy use, skill trajectory, social interactions
- Evidence Synthesis: May be simple or complex possibly using complex statistical models that may change over time

Items to Activities

Behrens & DiCerbo (2013) contrasted two ends of an assessment continuum as illustrated above by characterizing end points of an “Item Paradigm” and an “Activity Paradigm”. The Item Paradigm is associated with relatively focused tasks that are constrained to focus the scope of possible inferences from the observation. Typically, the task is also constrained to support scalable fixed response features such as multiple choice. DiCerbo and Behrens (2012) argue that this cost constraint was a major factor in the dominance of fixed response tasks (and thereby the item paradigm) during the digital desert. This also led to psychometric practices optimized on these practices and the corresponding constraint of the presentation processes to align with the restricted response scoring.

The activity paradigm starts with the assumption that in the new digital age, the facilities for presentation and evidence identification are not, and should not be, a primary constraint. By conceptualizing the assessment process as a feature extraction process from an activity (that may be constrained to fixed response but does not have to be), this conceptual model opens the possibility of assessment or general research data coming from a broad range of inputs including simulation based assessment (Frezzo, Behrens, Mislevy, West, & DiCerbo, 2009), online tutors (Feng & Heffernan, 2006), or other contexts that were perhaps not originally intended to serve assessment or instructional purposes (DiCerbo, in press).

	Item Paradigm	Activity Paradigm
Problem Formulation	Items pose questions	Activities request action
Output	Items have answers	Activities have features
Interpretation	Items indicate correctness	Activities provide attributes
Information	Items provide focused information	Activities provide multi-dimensional information

Table 1. Key differentiators between Item and Activity Paradigm from Behrens & DiCerbo (2013).

The conceptualization of the flexibility of the four process model is related to our ability to conceptualize and work in the activity paradigm. If we conceptualize the quantification process of measurement as a series of identifying specifically constrained answers (whether on a test, survey, or scoring rubric) then we have approached the problem with restricting limits to begin with and are likely to be driven increasingly toward the item paradigm. However, if we conceptualize the process as one of feature identification from a work product, then we have new, but less bounded problems. It does, however, free us up to extract new and often simultaneous observations from the activity data. We are freed to think of user activity as a complex stream from which we seek to observe certain attributes by applying observational rules over time, rather than a set of questions that should be scored for correctness. Of course, the second model is subsumed by the first. Taking this view opens up the possibility of complex scoring of activities in games (DiCerbo, in press; Valerie J. Shute & Ventura, in press), simulations (Frezzo et al. 2012), and ongoing system interaction across a number of attributes and behaviors, such as “gaming the system” (Baker et al., 2008).

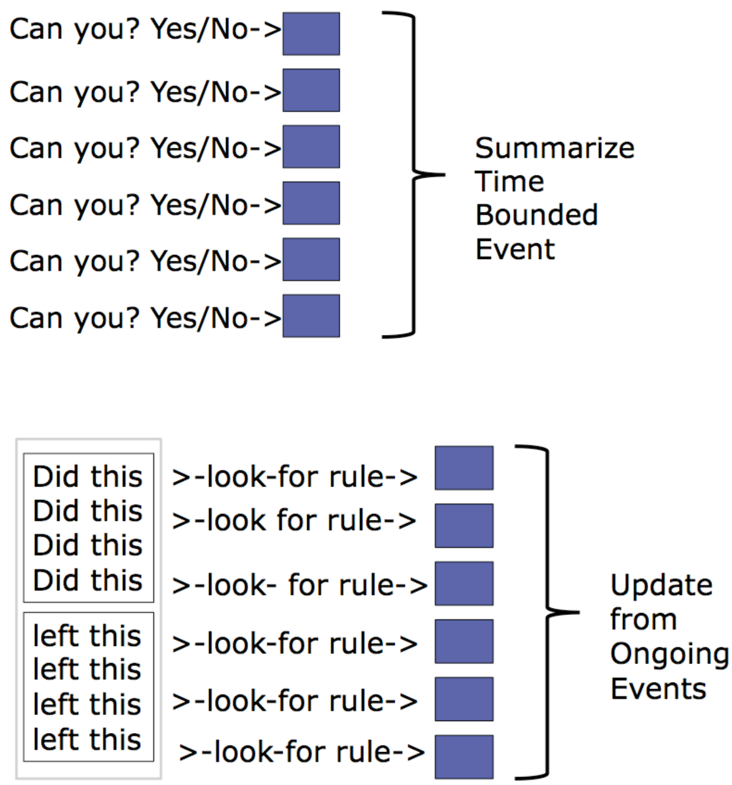

Figure 2: (a) characterization of the matching process in fixed response point-in-time assessment leading to summary scores, (b) characterization of generalized feature extraction process based on complex activity over time.

We may consider an additional extension of the conceptualization discussed by DiCerbo and Behrens (2012) with regards to the “presentation” process itself. Given their original context of assessment, using the term “presentation” and extending it to activity that “requests action” is an appropriate framining for that context. However, in attempting to expand the logic to a broader range of action, we may think not only about activities requesting action as in a test, but activities as interactions that invite, encourage, demand, attract, or otherwise motivate action and thought. To the degree assessment becomes based on observation of natural activity, the full range of purposes and contexts of activity, and the triggers of activity should be considered. Accordingly, the presentation process may be rightly renamed as an interaction or creation process given that the data-based and evidentiary outcome is a newly created work-product. Shifting this language from presentation (something the delivery system does) to interaction or creation (something the learner does) opens up new possibilities for metaphor and focus and sense-making regarding the activity of the learner. This view recommends a shift in the notion of features as fixed properties of tasks to features as emergent properties of interactions that may vary from individual to individual as different paths of action and creation provide different kinds of work products (play products? social products?) in complex systems. In the digital desert tasks and target features need to be highly constrained for evidentiary sense-making but in data rich environments forming the digital ocean, emergent features can be detected and combined in real time, as occurs in complex online gaming.

III. DATA SHIFT – UBIQUITY, CONNECTEDNESS, PERSISTENCE

The topics above have focused on the human activity that generates data to create the new digital ocean as well as the conceptual activity lens which we may use to understand the assessment/instruction/interaction process as it relates to data generation (presentation/interaction), transformation (evidence identification) and synthesis (evidence accumulation). In this section we discuss some of the issues related to the affordances of data storage. In this regard we discuss ubiquity, connectedness, and persistence, and contrast these new attributes of data based systems between digital desert and digital ocean scenarios.

Ubiquity

As envisioned by DiCerbo & Behrens (2012), the digital ocean exists because the shift to ever increasing natural interaction with sensor embedded devices allows the naturalistic and unobtrusive collection of data. In the digital desert, data collection was expensive, and dedicated resources and methods needed to be employed to collect and access the requisite data. In the digital ocean, data is being generated throughout the day by involvement with a myriad of systems. As those authors wrote

> "This is the vision of a world in which the natural instrumentation of a digital ocean blurred the distinctions between formative and summative assessment, curriculum and assessment, and formal and informal aspects of instruction. It 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." (DiCerbo & Behrens, 2012)

Insofar as the generation of data is device dependent, issues regarding appropriate methods for giving access need to be considered by the societies involved both with regard to the access to appropriate devices and the incentives and support for creating appropriate devices. Transformational technologies that are unable to be created for costs consistent with the economic issues of public education will fail to be adopted.

Inter-connectedness

To accomplish this goal, data collected from the sundry devices will need to be linked to be useful. Group level summaries of one system and group level summaries of another system fail to reveal the interactional effects that happen across variables and individuals. In the shorter term, individual systems will be built with internal linkages that preserve the user agreements and hopefully serve end users as desired by those end users. Because of the evolutionary nature of technology in education it is not uncommon for systems to be built separately for curricular or assessment data or formative and summative assessment systems. Systems designed this way fail to recognize the flexibility of the activity delivery framework and fail to take advantage of multi-dimensional linkages that may reveal important insights regarding patterns of learning.

Persistence

Persistence will be a third new and transformative characteristic of data in the age of the digital ocean. Persistence is important for several reasons. First, persistent data supports automated learning and decisions making from other systems (including human information gatherers). At present many aspects of the educational system are unaware of the user’s profile of previous experience. While human systems such as teachers may have developed detailed schema that are maintained and embellished over time, most activities systems start “cold” each time. For example, if a computer adaptive test (CAT) had a history of previous activity as a starting point in an assessment or tutoring activity, it may increase the efficiency of the assessment situation.

Second, persistent information may lead to improved knowledge models and research over time. Persistent information will also mean persistence of additional interpolated or computer generated attributes over time. Of course, the persistent nature of data raises many questions around privacy and data ownership, which unfortunately outstrip our current policies and regulations. These issues need to be addressed in reasonable ways that protect individuals, acknowledge the progress and potential of data availability while understanding the potential for harm.

IV. CORRESPONDING SHIFTS FOR RESEARCH AND TRAINING

The shift from digital desert to digital ocean will have dramatic impacts to educational research. We think these shifts will be dramatic and rapid and likely difficult to anticipate at present. Extrapolating from the observations discussed above, we suggest additional shifts that researchers and trainers of researchers should consider moving forward.

Theshift from data to answer questions to questions to answer data

In the digital desert the relationship between the research process and the data collection process is highly constrained by the cost of data. Present practice is to progress through a funnel of increasingly restricted ranges of data to be considered relevant as follows:

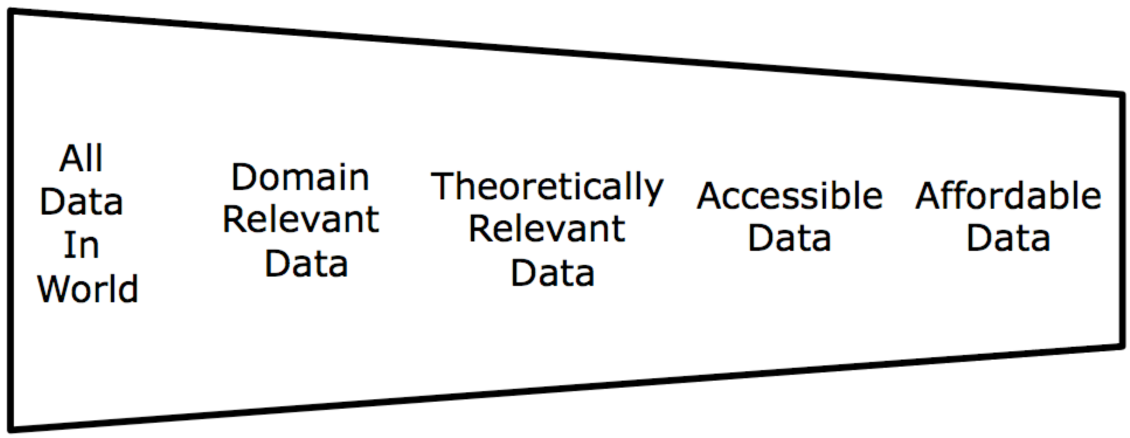

Figure 3: Funnel of data constraints on scientific data

While this analysis may appear cynical it likely explains the long-standing complaint that much institutional research is conducted on university undergraduates in laboratory conditions not because they are the most appropriate subject pool but rather because they are the most "affordable" data source given institutional support for the process.

As the digital ocean begins to rise and systems for many types of data availability begin to provide access to researchers, the types of questions which will be asked and how they will be addressed will change. We imagine this happening in at least two ways. First, easily-answered questions will be addressed and second, new forms of data and experience will create a theory gap between the dramatic increase in data-based results and the theory base to integrate them.

With regard to following easy data, we see this already in areas where large digital systems make some of the data available. For example, Twitter makes access to some of the data public on a daily basis. Computer tools for the extraction and visualization of this data are available to easily analyze and interpret some aspects of the data (Russell, 2011). Similarly Google (Google.com) provides online analytic tools to access results regarding search activity of their customers as well as basic word count analytics on their scanned text project.

While at present these explorations are likely justifiable endeavors, they raise the larger and long term issue of the role of data availability (and corresponding funding) in the scientific ecosystem. How much will the change in data availability privilege certain areas and how much will it open new areas of discussion?

Consider the filter model shown above in Figure 3, removal or dramatic release of constraints of the right side of the figure may move the setting of focal concerns toward broader theoretical settings. We see this for instance in the literature in educational assessment where some researchers have shifted focus from optimization of well established learning systems to new foci on motivation and intrinsic interest (e.g., Shute & Ventura, in press). Likewise, within the intelligent tutoring literature the availability of large amounts of learning data are now being complimented with sample-based data addressing more difficult human attributes (Baker et al., 2008). Supplementing the large automatically collected database with sample-based data collection from the classroom, these authors were able to address complex, long-term inferences in relatively efficient ways.

These shifts on the economic and conceptual constraints of data and theory availability have important implications for graduate training. They raise fundamental issues regarding the relationship between hypotheses, data, and analysis. We can easily imagine a world with a dearth of results seeking explanations as opposed to the current reverse situation of questions seeking data. In the context of scientific training and practice these shifts raise questions regarding the balance between ease of access to data and scientific value of addressing specific issues.

The shift to more results than answers

As the digital ocean evolves, it is likely that there will be times when there are more data to analyze than capacity to analyze it. In response, computationally sophisticated researchers will apply all-possible subsets of search strategies, for which theoretical explanations may, for a time, fall behind. Such activity is already evident in work undertaken using automated search and comparison over numerous data sets in the Pittsburgh Science of Learning Center's Data Shop open data and analysis infrastructure. New techniques such as Learning Factors Analysis (LFA; Cen, Koedinger, & Junker, 2006) attempts to recover series of optimally sloped learning curves across numerous combinations of possible variable combinations. While human guidance is often possible and likely preferred, large combinations of empirical results may be available compared with the number of available explanations.

While some may consider this an overly empiricalist approach, it appears at present as the natural automation of concepts and tasks currently undertaken in common educational and psychological inference in which hypothesis are often relatively underspecified leaving the room for a mix of conformational and "unexpected" results given a particular testing set up (Behrens, 1997; Gigerenzer, 2009). Moreover, with very large amounts of data over tens or hundreds of thousands of learners, there is like sufficient data for data exploration and hypothesis generation as well as confirmation on alternate data.

The Shift to more human interaction with digital devices

Human activity, both personal and social, will increasingly be facilitated by human interaction with electronic devices. Accordingly, students should have basic literacy in the understanding of Human Computer Interaction as a frame for research as well as literacy in modern software programming tools.

Methods for studying Human Computer Interaction have evolved greatly in the last 20 years along with the rise of human-machine interaction. While the moniker of "computer" is dated (as is "machine"), the idea that there are principles of interaction analysis can be brought to bear broadly in understanding human activity. For example, working from the Human Computer Interaction frames used in the Computer Supported Collaborative Learning literature to illustrate how Activity Theory (Engström, Miettinen, & Punamaki, 2007) can be applied to understand the human interactional dynamics of simulation based assessment and instruction. While this is often an embedded view in the learning sciences literature, it is not universally built into graduate study in education.

The shift from small computers for statistical or data collection to large systems for data collection and new analytics.

Following Siebel (2011), we believe that "Software is the new language of science". Understanding the logic of computing, the possibilities of modern applied computing and having facility for generic data manipulation and system interaction is essential. Recently the freely available and rapidly expanding Python language has emerged as a common tool for data visualization (Rossant, 2013; Vaingast, 2009), natural language processing (Bird, Klein, & Loper, 2009; Perkins, 2010), general data analysis (Janert, 2010; McKinney, 2012) and statistical manipulation (Conway & White, 2012; Russell, 2011). The R language is likewise emerging as a widely used tool for data science, though its statistical beginnings make it more appropriate for that arena than for solving universal computing problems.

Even if students are not going to obtain proficiency in a programming language, it is essential that they understand the basic logic of computing and trends in scientific computing. As a general overview and directions are needed for many researchers, this may be a ripe area for research supporting agencies to promote professional development.

Another way to help students prepare for the emerging digital ocean is to develop familiarity with standards for data description, movement, and use as embodied in standards for computer design and data exchange. For example, the Question and Testing Interoperability specification of the IMS (QTI; IMS, 2006) is an industry standard for assessment delivery. As a standard, it represents some level of consensus of practitioners in a field and represents the mental models prevalent at the time. Indeed, QTI was strongly influenced by the four process model described above, including specification of response processing and presentation processes. Other standards exist for other domains such as the predictive model markup language (PMML; Guazzelli, Lin, & Jena, 2012) used in data mining and related statistical disciplines.

As software tools become converging (even if only in discourse) points of activity across communities of educational practice, research, and development, it is increasingly important that training and research program address the current conceptualizations as represented in those systems.

The shift from research as event to research as an ongoing activity

Discussing the difference in grain size of data collection and feedback between digital desert and digital ocean Paradigms, DiCerbo & Behrens (2012) suggested the medical analogy as follows:

Educational Artifact	Medical Artifact
Summative End of Year Test	Autopsy
Formative Exam	Check up
Naturalistically Embedded Assessment	Heart Monitor

Table 2. Assessment granularity of educational artifacts and their corresponding analogs in medical examination.

One interesting implication of this model is that granularity of information is highly correlated with the sampling temporal frequency. This implies a shift from data collection as a series of isolated events triggered by causes unrelated to the phenomenon being studied to an ongoing interactional model of sensor/patient monitoring and engagement. The autopsy model supposes a drop in, point-in-time researcher (coroner) who is called to opportunistically take advantage of data collection opportunities. The heart monitor model assumes there is a responsible agent in partnership with the patient to both build agency in the activity and experience of the patient as well as to support and coach the patient on the basis of increased shared information.

The ubiquity and persistence of data represent additional complexity in the methodological landscape that has been traditionally dominated in educational research by time-agnostic or time-challenged methods such as simple Analysis of Variance or repeated measures analyses limited to a relatively few data points. New datasets that may contain hundreds or thousands of data points likely require new techniques to reflect the time and dimensionality complexities.

Likewise the shift in data granularity in the digital ocean open questions regarding whether the educational researcher plays the role of coroner or family doctor. Perhaps greater availability of data will allow the researcher to serve in a more active, continuous, supporting role while educators themselves become enabled by data access to become the newly empowered nurse practitioners. The determination of these outcomes will, in some part, depend on the evolving conceptual frames brought to the development of the devices and the human computer interactional features that evolve. It is incumbent on educational system designers to understand and study the implications of system design for learners and the stewards of learning (and research) in their ecosystems.

The shift from small sample research to large and combined data

The most prominent statistical frameworks of the last 100 years centered primarily around the problem of inferring population parameters from small samples (Behrens & Smith, 1996). Given the current move toward complete populations (see also Jager, Finite Sampling book) some common practice from these frameworks applied to large data can be misleading. For example, using a traditional significance test approach without considerations of effect size can actually increase inferential error (c.f. Glass, 1976). Accordingly, researchers are likely to need to be re-introduced to large sample or population analytic methods as the inferential value of digital desert methods recedes.

Shift from easy to handle data to hard to handle data.

The current Big Data movement (e.g., Franks, 2012) has often been defined less by the social/methodological implications discussed in this paper, but rather by the sheer size of the data and the necessity of developing new computing tools to address it (but see Smolan & Erwitt, (2012) for a compelling social view). For example, in large game data, a system may collect many millions of records of research data that cannot easily fit into individual machines or may extend the time required to complete and analysis to the point of making it untenable.

Students should be made aware of simple tools that can help resize and shape data. Tools such as SED and AWK and their derivatives allow for rapid extraction of key data from large files based on a simple query structure. Students will increasingly encounter Big Data that requires even more specialized approaches based on the specific technologies of Hadoop or other systems. In addition, advanced students should be familiar with the basic emerging algorithms that are becoming commonplace patterns in emerging computing. Computing for recommendation analysis based on collaborative filtering or other approaches seen in industry ("people like you bought X, you might want X"), for example, is an emerging common pattern (Ricci, Rokach, Shapira, & Kantor, 2010) that will eventually become part of the standard computing paradigm in education.

The shift from constrained and scored to open and computable

As noted in the second section above, we believe a key hallmark of the emerging digital ocean is that increase in open-form data that reflects the unstructured nature of human activity. This shift requires the acquisition and application of the conceptual tools discussed above in the context of the Four Processes Delivery model. These conceptual tools allow researchers to see beyond traditional data collection modes and give them a language around scientific discourse in educational domains.

In addition to the conceptual lens, student will also need to learn to compute and analyze data that is stream and event based. While this is an area of rich activity in some disciplines, advances with these types of data sources are only beginning to emerge in education (cf. Handbook of Educational Data Mining).

The shift from data new each time to ongoing model updating.

As discussed above, given the cost of data in the digital desert, research was often conducted at single points in time. The problem of lack of replication and publication bias exacerbate this concern and are well studied in the meta-analysis literature (cf. Hedges & Olkin, 1985). As standards for data collection, exchange and manipulation evolve and as access to ongoing-systems of data generation grow, there will be increased need for methodological approaches that not only describe the data at hand, but also provide an integration between new and existing data and information.

Two general approaches recommend themselves for this challenge: Bayesian statistics and meta-analysis. Meta-analysis was developed to address statistical weaknesses in the long-run and short-run probabilistic processes of significance testing and the lack of standardization in outcomes specification leading to difficulties in comparability (Glass, 1976). While meta-analysis is often conceptualized as a set of methods for summarizing a field or “conducting a literature review” the concepts and methods are amenable to many multi-results situations and recommend themselves even for analyzing results from within a single study, if the study is heterogeneous in its goals or data. A particularly valuable view of meta-analysis as a general analytic procedure is provided by Behrens & Robinson (2005) in which they suggested the importance of conceptualizing, analyzing, and displaying the results of multiple studies as a response surface reflecting the combined effects of the study characteristics.

Bayesian statistics have been designed precisely to take into account previously existing beliefs and conclusions and to provide a mathematical model for updating those conclusions. Accordingly, these approaches are well positioned to become a dominant paradigm in the digital ocean. In fact, the approach is currently experiencing an explosion of activity in the biological (Key & Schaub, 2012) and social sciences (Kruschke, 2012) because of the computational feasibility brought about by modern computing methods. Brooks, Gelman, Jones & Meng (2011) and Levy, Mislevy, & Behrens (2011) provide an overview of applications of Bayesian logic to education, and

Gelman & Hill (2006) provide an excellent account of social science methods building on both Bayesian and Frequentist ideas.

V. Concerns and Cautions

The notion of the digital ocean is not a proposal to increase instrumentation of learners for learning's sake. Rather it is attempting to give voice to the possibilities embedded in the social and technological shifts that are already occurring. Digital activity is becoming commonplace in daily life and it can change how we think about assessment, learning and education.

While detailed cautions and concerns can be enumerated at length for each section above, two broad concerns will need to suffice for this project.

First, the techno-social changes described in this paper and evidence around us are poorly understood as they relate to issues of social justice and equality. Differential access to devices or intelligent computation on one's data could lead undesirable social outcomes as new types of under-served roles evolve. Educational economists and policy experts should be involved in the conversations regarding the implications of these changes for local, national and global socio-political systems.

Second, with regard to the academic/scientific communities, it is fitting to review the stories recounted in Stephen Jay Gould's *The Mismeasure of Man*. While there was some controversy and disagreement over his characterization of some relatively recent scholars (e.g., disagreements regarding his characterizations of Arthur Jenson), the long historical view painted a portrait of "modern" science that requires no advanced degree to raise concern. In this volume, Gould repeatedly recounts leading scientific experts of the 19th and 20th centuries coming to socially destructive and biased conclusions on the basis of new and irrefutable use of scientific data. These "objective" scientific conclusions in which the data were "allowed to speak" led to policies including mass sterilization and support for Nazi eugenics. It seems an appropriate time to review Gould's stories and engage the philosophy and history of science communities in dialogs regarding how to most appropriately harness the data from this ever changing world.

We are at the very dawn of a great intellectual revolution. A great Renaissance or perhaps more appropriate, a great Enlightenment in which not only do we do things differently, but the strength of the difference is palpable in the historic mind, and the new experiences cause us to reflect on the fundamental issues of our past endeavors. Or perhaps we are at the start of another great Post-Industrial revolution in which the nature of previous social fabrics is changed in light of the economics of human systems.

Regardless of the appropriate frame, all our experience in the last 19 years, since the introduction of the World Wide Web allowed for near universal communication between humans as well as the movement of data and computational results among machines, suggests that the tide of data is rising dramatically and that new conceptualizations are needed to understand both our past and future relationships with data, analysis, and each other.

Then felt I like some watcher of the skies
  When a new planet swims into his ken;
Or like stout Cortez, when with eagle eyes
  He stared at the Pacific—and all his men
Look'd at each other with a wild surmise—
  Silent, upon a peak in Darien.

J. Keats,

Final stanza’s from “On first looking into Chapman's Homer”

Author Notes:

John T. Behrens is Vice President and leader of the Center for Digital Data, Analytics & Adaptive Learning at Pearson. He is also an Adjunct Assistant Research Professor in the Department of Psychology at the University of Notre Dame (Indiana, USA) and would like to thank the department for its support.

He would also like to thank Drs. Kristen E. DiCerbo, Robert J. Mislevy, and Philip Piety for logical, psychological, and textual support of this paper and related activity Thanks also to Quinn Lathrop and Shauna Sweet for providing helpful comments on a draft of this paper. All errors and weaknesses are mine.

References

Almond, R. G., DiBello, L. V., Moulder, B., & Zapata-Rivera, J.-D. (2007). Modeling diagnostic assessments with Bayesian networks. *Journal of Educational Measurement*, *44*(4), 341–359.

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 https://escholarship.bc.edu/ojs/index.php/jtla/article/view/1671

Baker, R., Walonoski, J., Heffernan, N., Roll, I., Corbett, A., & Koedinger, K. (2008). Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, *19*(2), 185–224.

Behrens, J. T. (1997). Principles and procedures of Exploratory Data Analysis. Psychological Methods, 2, 131-160.

Behrens, J. T., & Robinson, D. H. (2005). The micro and the macro in the analysis and conceptualization of experimental data. In G. D. Phye, D. H. Robinson, & J. Levin (Eds.), Empirical Methods for evaluating educational interventions.(pp. 147-173) Burlington, MA: Elsevier.

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: Volume 2, Research Methods in Psychology* (2nd Edition). (pp. 34-70). New York: Wiley.

Behrens, J. T., Frezzo, D., Mislevy, R., Kroopnick, M., & Wise, D. (2006). Structural, functional and semiotic symmetries in simulation-based games and assessments. *Assessment of problem solving using simulations*, 59–80.

Behrens, J. T., & DiCerbo, K. E. (2013). *Technological Implications for Assessment Ecosystems: Opportunities for Digital Technology to Advance Assessment*. Princeton, NJ: The Gordon Commission on the Future of Assessment.

Behrens, J. T., & Smith, M. L. (1996). *Data and data analysis.*

Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python* (1st ed.). O'Reilly Media.

Cen, H., Koedinger, K., & Junker, B. (2006). Learning factors analysis–a general method for cognitive model evaluation and improvement. In *Intelligent Tutoring Systems* (pp. 164–175). Retrieved from http://link.springer.com/chapter/10.1007/11774303_17

Conway, D., & White, J. M. (2012). *Machine Learning for Hackers* (1st ed.). O'Reilly Media.

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 http://www-jime.open.ac.uk/jime/article/viewArticle/2012-18/html

DiCerbo, K. (in press). *Game-based Measurement of Persistence*.

DiCerbo, K., & 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* (pp. 273–306). Charlotte, NC: Information Age.

Engeström, Y., Miettinen, R., & Punamäki, R.-L. (1999). *Perspectives on activity theory*. Cambridge University Press

Feng, M., & Heffernan, N. T. (2006). Informing teachers live about student learning: Reporting in the assistment system. *Technology Instruction Cognition and Learning*, *3*(1/2), 63.

Frezzo, D. C., Behrens, J. T., Mislevy, R. J., West, P., & DiCerbo, K. E. (2009). Psychometric and evidentiary approaches to simulation assessment in Packet Tracer software. In *Networking and Services, 2009. ICNS'09. Fifth International Conference on* (pp. 555–560). Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4976818

Gee, J. P. (2003). What video games have to teach us about learning and literacy. *Computers in Entertainment (CIE),* 1(1), 20–20.

Gelman, A., & Hill, J. (2006). *Data Analysis Using Regression and Multilevel/Hierarchical Models* (1st ed.). Cambridge University Press.

Glass, G. V (1976). Primary, secondary, and meta-analysis of research. *Educational* researcher, 5(10), 3–8.

Guazzelli, A., Lin, W.-C., & Jena, T. (2012). *PMML in Action: Unleashing the Power of Open Standards for Data Mining and Predictive Analytics* (2nd ed.). CreateSpace Independent Publishing Platform.

Hedges, L. V., & Olkin, I. (1985). *Statistical Methods for Meta-Analysis* (1st ed.). Academic Press.

IMS. (2006). IMS question & test interoperability specification v2.0/v2.1. <http://www.imsglobal.org/question/index.html> Retrieved March, 2013.

Janert, P. K. (2010). *Data Analysis with Open Source Tools* (1st ed.). O'Reilly Media.

Kappiarukudil, K. J., & Ramesh, M. V. (2010). Real-Time Monitoring and Detection of "Heart Attack" Using Wireless Sensor Networks. In 2010 Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM) (pp. 632–636). Presented at the 2010 Fourth International Conference on Sensor Technologies and Applications (SENSORCOMM). doi:10.1109/SENSORCOMM.2010.99

Kery, M., & Schaub, M. (2011). *Bayesian Population Analysis using WinBUGS: A hierarchical perspective* (1st ed.). Academic Press.

Kruschke, J. K. (2010). *Doing Bayesian Data Analysis: A Tutorial with R and BUGS* (1st ed.). Academic Press.

Levy, R., Mislevy, R.J.,& Behrens, J.T. (2011). MCMC in educational research. In S. Brooks, A. Gelman, G. Jones, G., & X. L. Meng (Eds.). *Handbook of Markov Chain Monte Carlo* (pp. 531-546). Boca Raton: Chapman and Hall/CRC.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). An introduction to latent semantic analysis. *Discourse processes*, 25(2-3), 259–284.

McKinney, W. (2012). *Python for Data Analysis.* O'Reilly Media.

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

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

Pearl, J. (1988). *Probabilistic Reasoning in Intelligent Systems: Networks of Plausble Inference*. Morgan Kaufmann Pub. Retrieved from http://books.google.com/books?hl=en&lr=&id=AvNID7LyMusC&oi=fnd&pg=PA1&dq=Pearl+bayesian+networks&ots=FY-OSfkwZ6&sig=3J1ZLPeMMUfZNG_k73CkHClSj7o

Perkins, J. (2010). *Python Text Processing with NLTK 2.0 Cookbook.* Packt Publishing.

Ricci, F., Rokach, L., Shapira, B., & Kantor, P. B. (Eds.). (2010). *Recommender Systems Handbook* (2011th ed.). Springer.

Rossant, C. (2013). *Learning IPython for Interactive Computing and Data Visualization.* Packt Publishing.

Russell, M. A. (2011). *Mining the Social Web: Analyzing Data from Facebook, Twitter, LinkedIn, and Other Social Media Sites* (1st ed.). O'Reilly Media.

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

Shute, V. J. (2011). *Stealth assessment in computer-based games to support learning. Computer games and instruction.* Charlotte, NC: Information Age Publishers. Retrieved from http://myweb.fsu.edu/vshute/pdf/shute%20pres_h.pdf

Shute, Valerie J., & Ventura, M. (in press). *Measuring and supporting learning in games: Stealth assessment*. 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), 30–32.

Smolan, R., & Erwitt, J. (2012). The Human Face of Big Data (First Edition.). Against All Odds Productions.

Vaingast, S. (2009). *Beginning Python Visualization: Crafting Visual Transformation Scripts* (1st ed.). Apress.

Van der Linden, W. J., & Hambleton, R. K. (1996). *Handbook of modern item response theory*. Springer. Retrieved from http://books.google.com/books?hl=en&lr=&id=aytUuwl4ku0C&oi=fnd&pg=PR5&dq=hambleton+item+response+theory&ots=JXdX5GjwfM&sig=KKbXwfRqSFqzMLsU_2BTY-rhcFk

Wolf, G., Carmichael, A., & Kelly, K. (2010). The quantified self. *TED http://www. ted. com/talks/gary_wolf_the_quantified_self. html*.

Wolf, Gary. (2002, October 5). The Data Driven Life. *The New York Times Magazine*, 38–45.

Last Revision June 5, 2013