Title of Document:

Research on Implementing Big Data:

Technology, People, & Processes

Author Names:

Jenny Grant Rankin, Ph.D.; Margie Johnson, Ed.D.; and Randall Dennis

Conference Organization:

Society for Information Technology & Teacher Education (SITE)

Conference Title:

Society for Information Technology & Teacher Education (SITE) International Conference

Conference Location: Las Vegas, Nevada

Date: March 3, 2015

The subsequent pages comprise the complete presentation paper

as it was submitted, accepted, and presented.

Research on Implementing Big Data: Technology, People, & Processes

Jenny Grant Rankin, Ph.D.
Formerly Illuminate Education
United States
drjrankin@gmail.com

Margie Johnson, Ed.D.
Metro Nashville Public School
United States
margie.johnson@mnps.org

Randall Dennis
Intelligex
United States
randall.dennis@intelligex.com

Abstract: When many people hear the term "big data", they primarily think of a technology tool for the collection and reporting of data of high variety, volume, and velocity. However, the complexity of big data is not only the technology, but the supporting processes, policies, and people supporting it. This paper was written by three experts to address all parts of a big data system: technology, processes, and people. Randall Dennis is Chief Strategy Officer for an education data analytics firm listed three times on the INC 5000, founder of a strategic consulting firm, and is directly involved in education data analytics product development. Dr. Jenny Rankin has been a teacher, school administrator, district administrator, and Chief Education & Research Officer of Illuminate Education. Dr. Margie Johnson has been the Business Intelligence Coordinator for Metro Nashville Public Schools for three years and empowers approximately 10,000 employees in data-informed decision-making.

Introduction

When many people hear the term "big data" they primarily think of a technology tool for reporting on enterprise-scale data. However, *big data* refers not only to scale of data and tools, but to a complex system made up of not only the technology, but also the processes and people involved in data collection, analysis, and use. Just like a symphony, there are three basic components to a big data system: technology (instruments), processes (musical scores), and people (the performers). Our panel discusses these components in depth:

- **Technology** – Data systems involve three general phases: collection, management, and utilization. Each phase is supported by different technologies. Challenges include data integrity, aggregation, and alignment. The great opportunities relate to new tools to leverage data for analysis and data visualization for high-stakes decision making.

- **Processes** – Implementing big data provides a spotlight for data issues within an organization. Because departments were used to working in silos, many times different processes are used to collect the same data. When consolidating big data, data governance is needed in an organization. Data governance refers to the rules, decision rights, and accountabilities (i.e. processes) of people and technology as they perform data-related processes (Data Governance Institute, 2007). With data governance, an organization institutes processes to understand the cause and effect of poor data, so that solutions can be developed to correct the problem and a means for monitoring and evaluating the implementation of these solutions can be adopted.

- **People** – Data is an organizational asset just as the people who work for an organization. Therefore, Big data is most effective when it fosters business intelligence throughout the organization and helps build the

capacity of educators to make informed decisions. A data-informed decision making framework developed from research and practice at the Metropolitan Nashville Public Schools will be shared.

Technology

Data systems involve three vital phases: collection, management and utilization. Each phase is supported by different technological tools. Among big data's greatest challenges to data-informed decision-making are a.) ensuring supporting data technologies ensure high integrity, b.) effective aggregation and alignment of various disparate datasets in varying formats, and c.) leveraging useful tools to analyze and visualize data to make effective data-informed decisions. The session will explore some of the technologies associated with each of the three phases of big data projects -- and the inherent strengths and weaknesses of these tools. Some tools are general open source tools requiring expert customization are available, while new education-specific solutions are emerging.

The session will review the strengths and weaknesses of a number of technology resources for education data acquisition (including xAPI specifications such as Tin-Can API), warehousing, and analytics tools, whether general tools (such as Apache Hadoop, Tableau and Roambi) or education specific tools (Knewton, RANDA empower, etc.). Further, we'll discuss some of the inherent perils associated with data relating to utilizing student data (identification, demographics, summative & formative assessments, course completion, post-secondary readiness) teacher (identification, demographics, observations, PD/CE), school (climate, etc.) and third party data (census, crime maps, etc.) tools in the context of Dennis' "Enterprise Education Data Confidence Model".

Processes and Policies

In schools and school districts, processes should be in place to ensure data is properly collected, inputted, maintained, reported, analyzed, and used. These stages in data's journey also relate to the technology and people involved with education data, and thus these themes are also echoed in this paper's other sections. Related processes are integrated within Over-the-Counter Data Standards (www.overthecounterdata.com/s/OTCDStandards.pdf), which are education data reporting standards where data usage support is embedded within the usage environment, as is the case with over-the-counter products. The standards are based on over 300 studies and other expert sources related to best practices for communicating data to educators. Consciously adhering to process standards and guidelines, rather than leaving staff to navigate data stages on their own, will best facilitate success for staff and students when the data is ultimately used.

Data Collection

Thousands of data elements are collected in an education data system for reporting purposes (Colorado Department of Education, 2008). For example, 303 data types can be seen in Sample Data Types to Support Educators' Data Analyses (www.overthecounterdata.com/s/DataTable.pdf). This document captures the variety of data collected and demonstrates that educators need to be collecting and using more than simply assessment data.

Careful consideration should be paid to any data collection tools prior to the acquisition of actual data. For example, assessments and other performance measurements should be tightly aligned with what is being taught and when it is being taught, and assessment quality should be evaluated and attained. Even after these tools are used, data should be leveraged to identify any problem areas, such as misleading test questions (e.g., noticed when an overwhelming number of students select a particular distractor), or tasks that are poorly aligned with the standards they are meant to assess (e.g., noticed when results do not inform the part of the standard with which students struggled). Additional standards that can inform use of data collection instruments are the Code of Professional Responsibilities in Educational Measurement (National Council on Measurement in Education, 1995); American Educational Research Association (AERA) Standards for Educational and Psychological Testing (AERA, American Psychological Association, & National Council on Measurement in Education, 1999 version); and Code of Fair Testing Practices in Education Reporting and Interpreting Test Results (Joint Committee on Testing Practices, 2004).

Data collection processes should be clearly communicated to staff. For example, before a common assessment is administered, all staff should know relevant administration windows and guidelines for test preparation, administration, and score submission. These processes should evolve when staff feedback warrants change. For example, guidelines on acceptable study tools might warrant revision after some teachers are found giving students copies of the test to use as study guides. Data collection should also be tightly aligned with data access so staff can use data immediately, or else as close to immediately as possible.

Data Governance

Data privacy and processes should adhere to the Family Educational Rights and Privacy Act (FERPA) throughout all data stages, beginning with data collection. For example, collected parent surveys should not be placed anywhere that is not secure, as they should only be accessible to those whose access is warranted. However, data privacy is most commonly associated with the entry and maintenance of data, which are part of data governance. Though data can only be managed after it has been collected, data governance is an ongoing stage that involves ensuring data is accurate, clean (e.g., no duplicate records), comprehensive (e.g., no missing records), accessible to stakeholders, and secure. Clear processes are required for these qualities to be achieved.

Data quality is paramount to effective data use. All staff responsible for entering and/or managing data should have access to a district-wide data input matrix to guide them in understanding data system fields so data is inputted appropriately. For example, front office staff can use the matrix when fielding registration forms or entering demographic changes, and teachers can use the matrix if they are tasked with using data system information to complete non-computerized testing answer documents that missed pre-ID. This matrix can be accompanied by details on which staff is responsible for which stages of data entry, such as ensuring original data files are appropriately formatted and complete, uploading files, and more. These details can vary by dataset and include contact information so other stakeholders know who to contact if data errors are found within a data system. All data governance processes should be clearly communicated to staff, evolve as necessary, and accompany regular, internal audits for data quality.

The Metropolitan Nashville Public Schools (MNPS) has made significant investments in data systems and supports to facilitate data informed decision making from the classroom to executive leaders. Historically, data was collected and managed at the level of individual departments for their own needs. Each department has developed procedures, data formats, and terminology (i.e., processes) that address its unique situation and preferences. As long as there was no need to integrate or exchange the data, such inconsistencies were harmless.

Today, MNPS' strategic plan, mission and legal mandates require MNPS to report on our activities at the enterprise level. This means that MNPS needs to:

- Migrate data from existing systems into new systems and formats.
- Integrate and synchronize data from different systems that use different formats, field names, and data characteristics.
- Reconcile inconsistent or redundant terminology through a single data dictionary providing agreed upon definitions and properties for each data element
- Manage metadata with the purpose of facilitating the discovery of relevant information, organizing electronic resources, and supporting the archiving and preservation of data.
- Report data in standard formats and with standard interpretations.

Data governance (Fig. 1) helps MNPS implement big data by providing and enforcing enterprise-wide data standards, common vocabulary, reports, and the development and use of standardized data. It enables MNPS to more easily integrate, synchronize and consolidate data from different departments, exchange data with other organizations in a common format, and communicate effectively through shared terms and report formats.
Data governance is a program which addresses data throughout the district. Support for the program begins at the executive level and continues through all departments and all employees. As data is created, stored, and used at all levels of the organization, data governance encompasses opportunities and responsibilities at all levels as well.

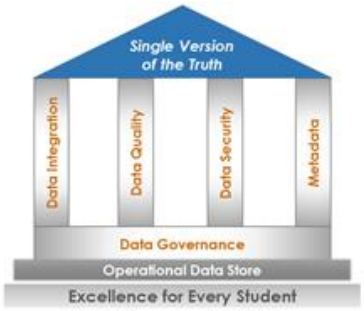

Figure 1: Data Governance

Data governance puts personnel, policies, procedures, and organization structures in place to make data accurate, consistent, secure, and available to accomplish MNPS' mission. Data governance provides district-wide data standards, common vocabulary, reports, and the development and use of standardized data. With data governance, MNPS employees are empowered to access and manage data assets with their assigned responsibilities. The ultimate goal of a data governance strategy is to make MNPS more efficient by saving money, allowing re-use of data, supporting data-informed decision making, and to treat data as the asset it is to the district.
Data governance refers to the rules, decision rights, and accountabilities of people and technology as they perform data-related processes (Data Governance Institute, 2014). Data governance ensures that data can be trusted and supports the improvement of data quality by identifying root causes of data issues. It is about putting people in charge of correcting and preventing data issues in order to maximize the impact of our data assets. The goals of data governance are to make information:

- Reliable
- Consistent
- Complete
- Easily available to those with a legitimate need for it
- Unavailable to those without a legitimate or authorized need for it

People

There is a great deal of interest in the idea of "big data," and the power big data represents for organizations. What is often lost in the discussion is that having a big data system is only one component of using big data. Big data is most effective when it fosters business intelligence throughout the organization and helps build the capacity of educators to make informed decisions. In her work as Business Intelligence Coordinator for the 47th largest school district in the United States, Dr. Margie Johnson has translated the latest research, including Dr. Jenny Rankin's work, on data use into a data-informed decision making framework (Fig. 2).

Figure 2: Data-Informed Decision Making Framework

Culture of Collaborative Inquiry

You may have heard the saying that we are "data rich, but information poor." Having access to data does not change the way we work. Instead, a data-informed organization must be committed to developing a culture of collaborative inquiry, which provides structures and processes for how an organization can work together to ask questions, solve problems, and ask more questions to support high student achievement. By using this approach, the data may change on a regular basis based on the questions being asked, but the way of working together does not.

Common Language

Developing common language around data-informed decision making initiatives is critical. If existing language exists that can be leveraged, then use it to create a common language of continuous improvement and collaborative inquiry. Whenever possible, do not reinvent the wheel. For example, in Dr. Johnson's organization, the 8-Step Continuous Improvement Model (CIM) by Dr. Patricia Davenport existed. Therefore, that model was communicated throughout the district as the collaborative way of using data. Teachers were provided with common planning time and CIM provided a basic framework for structuring the conversations.

Data Access

Since data are everywhere, accessing data can sometimes become a laborious process. MNPS invested Race to the Top funds to bring down the data silos that exist in most organizations and build one of the most robust data warehouses in the United States. Not only is the data consolidated, but educators throughout the organization have access to it. Furthermore, the MNPS Data Warehouse team can build adapt and customized reports as requested depending on the questions that need to be answered.

Data Literacy and Analysis

The National Center for Education Statistics estimates less than 2% of school districts in the U.S. are able to turn the data languishing in data warehouses into information educators can use (Sparks, 2014). There is abundant evidence educators of all levels have trouble interpreting data (Underwood, Zapata-Rivera, & VanWinkle, 2010). These studies include the Rankin (2013) quantitative study, in which 211 educators with varied backgrounds at nine schools throughout California analyzed data within varied environments. Educators' data analyses were shown to be 11% correct when using typical data reports. However, this accuracy rose by up to 436% when they received data in an over-the-counter format, meaning support for understanding the data was embedded within the reporting environment.

Educators are largely not to blame for their data analysis errors. Rather, data is often not presented in the format educators need to use the data properly (Underwood et al., 2010). Educator leaders should advocate for their data systems and reports to adhere to Over-the-Counter Data Standards to ensure data is presented in ways educators can easily understand. This involves appropriate data visualization and accuracy, but it also involves offering data analysis support within the tools educators use to view data. For example, as discovered in the Rankin (2013) study:

- Educators' data analyses were 307% more accurate when data reports featured a footer offering guidance in the data's meaning, and 336% more accurate when respondents reported using the footer.
- Educators' data analyses were 205% more accurate when data reports were accompanied by a 1-page reference sheet offering help understanding the report's data, and 300% more accurate when respondents reported using the reference sheet.
- Educators' data analyses were 273% more accurate when data reports were accompanied by a 2- to 3-page reference guide offering help understanding and using the data report, and 436% more accurate when respondents reported using the reference guide.
- A shorter, targeted manual or user-friendly help system caused users to need 40% less training time and to successfully complete 50% more tasks than would be accomplished with only access to a full-sized manual (van der Meij, 2008).

MNPS uses two strategies to build improve educators' data literacy and analysis skills—data guides and data coaches. Dr. Jenny Rankin conducted a study (2013) entitled, *Over-the-Counter Data*'s *Impact on Educators' Data Analysis Accuracy*, which revealed that educators' data analysis skills was up to 307% more accurate when data supports, including headers, footers and data guides were embedded in the data system.

For student achievement to improve, teachers need to learn to transfer newly gained knowledge and skills into practice (Nolan & Hoover, 2008). In 2002, a meta-analysis of 200 research studies compared the relationship among the training components included in the professional development and the attainment of three outcomes categories: knowledge, skill demonstration, and use in the classroom. The highest transference occurred when teachers participated in training but also received follow-up coaching in the classroom, which resulted in a 95% implementation rate in all three outcome categories (Joyce & Showers, 2002, p. 78). MNPS implemented this high yield professional development strategy by hiring district data coaches.

Conclusion

Orchestrating an effective education big data initiative is as complex undertaking. Instruments (technology), scores (processes) and people (the players) must work in concert to achieve a harmonious. Whether your role is the conductor, the composer or a player, big data initiatives rise or fall based on attention to detail in all three.

References

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (1999). *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association.

Colorado Department of Education. (2008). *Statewide Longitudinal Data Systems (SLDS) grant: Progress overview, July 2008*. Retrieved from http://www.cde.state.co.us/sites/default/files/documents/slds/download/pdf/slds_progressoverview_as_of_july_2008.pdf

Data Governance Institute. (2014). Definitions of data

Joint Committee on Testing Practices (2004). *Code of fair testing practices in education*. Washington, DC: American Psychological Association. Retrieved from on http://www.apa.org/science/programs/testing/fair-code.aspx

National Council on Measurement in Education (1995). *Code of professional responsibilities in educational measurement*. Washington, DC: Author.

Rankin, J. G. (2013). *Over-the-counter data's impact on educators' data analysis accuracy*. ProQuest Dissertations and Theses, 3575082. Retrieved from http://pqdtopen.proquest.com/doc/1459258514.html?FMT=ABS

Sparks, S. D. (2014, July 25). Can states make student data useful for schools? *Education Week*. Retrieved from http://blogs.edweek.org/edweek/inside-school-research/2014/07/can_states_turn_slag_data_into.html

Underwood, J. S., Zapata-Rivera, D., & VanWinkle, W. (2010). *An evidence-centered approach to using assessment data for policymakers* (ETS Research Rep. No. RR-10-03). Princeton, NJ: ETS.

Van der Meij, H. (2008). Designing for user cognition and affect in a manual. Should there be special support for the latter? *Learning & Instruction*, 18(1), 18-29.