Data competence maturity: developing data-driven decision making

Thomas G. Cech

National Council for Community and Education Partnerships, Washington, District of Columbia, USA

Trent J. Spaulding

Department of Health Care Management, Appalachian State University, Boone, North Carolina, USA, and

Joseph A. Cazier

Department of Computer Information Systems, Appalachian State University, Boone, North Carolina, USA

Abstract

Purpose – The purpose of this paper is to lay out the data competence maturity model (DCMM) and discuss how the application of the model can serve as a foundation for a measured and deliberate use of data in secondary education.

Design/methodology/approach – Although the model is new, its implications, and its application are derived from key findings and best practices from the software development, data analytics and secondary education performance literature. These principles can guide educators to better manage student and operational outcomes. This work builds and applies the DCMM model to secondary education.

Findings – The conceptual model reveals significant opportunities to improve data-driven decision making in schools and local education agencies (LEAs). Moving past the first and second stages of the data competency maturity model should allow educators to better incorporate data into the regular decision-making process.

Practical implications – Moving up the DCMM to better integrate data into their decision-making process has the potential to produce profound improvements for schools and LEAs. Data science is about making better decisions. Understanding the path laid out in the DCMM to helping an organization move to a more mature data-driven decision-making process will help improve both student and operational outcomes.

Originality/value – This paper brings a new concept, the DCMM, to the educational literature and discusses how these principles can be applied to improve decision making by integrating them into their decision-making process and trying to help the organization mature within this framework.

Keywords Data, Secondary education, Analytics, Data competency maturity model, Data-driven decision making

Paper type Conceptual paper

Introduction

Over a decade of research has called for better use of data in education (Honig and Venkateswaran, 2012; Reeves, 2017), yet most schools and local education agencies (LEAs) still struggle to fully use their data to make better decisions (Slavin *et al.*, 2013; Grissom *et al.*, 2017). Many types of data are flooding secondary education (Horn *et al.*, 2015). Organizational and political issues as well as a haphazard approach to data storage have

hindered the use of data to improve school performance and student and teacher experiences. A competency maturity model has the potential to guide schools and LEAs through developing the ability to use the data they collect. This essay develops a data competency maturity model to lay out a theoretical and practical path to leverage data in secondary education. The purpose of this model is to allow educators to better leverage the data they have today and to begin to plan how to use the data they may be able to get in the future to make better decisions.

More mature practices around the use of data enable educators to extract the most value from the data. Data analytic techniques have the potential to improve school performance by providing teachers, administrators, advisors and counselors (collectively referred to as educators hereafter) with more evidence for decision making and improved early warning systems. Data are captured to ensure compliance with state and federal requirements and then used for various annual report cards at the student, school, state and federal levels (Grissom *et al.*, 2017). These data are often used to describe outcomes (e.g. which students do not graduate, which students are at risk of dropping out and what schools can do to prevent dropping out (Lai and Mcnaughton, 2016)); however, more value is available in the discovery of relationships and causes.

Our data are stored in a variety of structured and unstructured formats (e.g. file cabinets, desktops, laptops, servers and on the cloud through vendor solutions). Some data are readily available (e.g. attendance, grades, extracurricular activity, online behavior, social networks information, etc.), and other data often take more effort to obtain (home environments, income levels, parental education, parental involvement, etc.). Data sources are often disjointed and not immediately available to those who need it. Even when data are available, decision makers are often unaware of the data and do not have the tools or skill set necessary to leverage the data (Weinbaum, 2015; Hubers *et al.*, 2017). When we use both the educator's knowledge and all available data, we are more likely to achieve better outcomes for students (Warner, 2014).

Currently, secondary and post-secondary institutions are using analytics to improve their services and for improving various key performance indicators (e.g. grades, retention). Presently, two analytic techniques (educational data mining and learning analytics) are used to experiment and develop models in several areas that can improve learning systems and educator/school performance (Bienkowski *et al.*, 2012). Where educational data mining (e.g. mining website log data to build models) tends to focus on developing new tools for discovering patterns in data, learning analytics (e.g. prescribing interventions to students identified as at-risk) focuses on applying tools and techniques at larger scales. Both analytic techniques work with patterns and prediction and, if applied correctly, have the potential to shed light on data that have not been acted on. The unique competency maturity model developed and presented in this work puts these and other practices in the perspective of appropriate development of analytic capabilities. The theoretical and conceptual foundation laid for the data competency maturity model here should provide opportunities to better assess data usage and develop data capabilities in secondary education. Further, this new approach to maturity in data analytics may have implications in many other fields where data analytics is still maturing. The model developed here will need further empirical and practical validation.

Literature review
Applications of data analyses
The purpose of applying analytics practices in education is to provide tools to make the best and most reliable decisions. Making the best decisions requires optimally matching the analysis to the question. Extant literature regarding data and evidence-based decision making classify the types of questions into four categories: descriptive, diagnostic,

predictive and prescriptive (Conner and Vincent, 1970; Banerjee *et al.*, 2014; Yadav and Kumar, 2015). In the literature, these categories are often referred to as analyses or applications of analyses. Nevertheless, it is critical in this work to make a distinction between application and analysis. For clarity, we refer to these applications as questions or the types of questions being answered.

Descriptive questions focus on what has happened in the past and what is currently happening. Answering descriptive questions does not involve specifying relationships in the data. For example, educators need to know what percent of students graduated, how many students passed standardized tests and how their average test scores compare against other schools.

Diagnostic questions attempt to explain why a particular event has occurred (Wilson and Demers, 2014). Correlations, paired with solid theory, can suggest that there is a relationship between two factors. An administrator could use the findings of an analysis to list the factors related to GPA (e.g. attendance). However, understanding why something happens is different than implementing an intervention to affect the outcome.

Predictive questions are concerned with predicting what is likely to happen in the future (Milliken, 2014). Clearly answering predictive questions requires models to identify patterns in the available data. The predictive application of data has been successfully employed in a number of industries and organizations (Davenport and Harris, 2007). If school administrators can better understand their students and teachers, then they can anticipate needs and make operational improvements based on likely future events. Jindal and Borah (2015) lay out a case for the use of predictive analytics in education.

Prescriptive questions attempt to determine if a specific intervention will have a specific outcome. Prescriptive questions can also employ predictive models as well as optimization techniques to recommend one or more courses of action (Koch, 2015). Prescriptive questions attempt to intervene to achieve the desired outcomes. Avoiding unintended consequences in answering prescriptive questions requires great care in forming and using an appropriate analysis.

Types of data analyses

Four classifications of analyses are emphasized in the literature. They include non-empirical (Romanczyk *et al.*, 1973; Lowin and Craig, 1968), summary (Trochim, 2000), correlational (Cliff, 1983; Trochim, 2000) and causal analyses (Cliff, 1983; Trochim, 2000).

Non-empirical analysis is typically derived from casual observations rather than from experiments or formal data collection (Cotton *et al.*, 2010). Information gathered through common, unstructured observation is often employed to make decisions (Hubers *et al.*, 2017). A teacher may conclude from the (mis)behavior of the class that the students are tired (Horn *et al.*, 2015). Non-empirical analysis provides valuable insight, particularly when data are not available or it is the only type of analysis available to an educator. These types of judgments are often heavily relied upon in practice (Coburn and Talbert, 2006). However, this type of analysis cannot guarantee that one phenomenon is causing another. Human judgment can be biased.

Summary analysis is the most basic form of quantitative analysis and is commonly used in teacher evaluation, student performance evaluation and school performance. Summary analysis can also be referred to as descriptive analysis. This type of analysis uses statistical measures such as means, standard deviations and percentiles. It is much more reliable than non-empirical analysis, particularly in determining differences between groups or what constitutes an outlier.

Correlational analysis investigates a statistical relationship between two phenomena. Correlational analysis begins the journey to understanding the "why" of an issue, but it does not state that one variable causes the other. Most educators understand there is a relationship

between attendance and GPA (Allensworth *et al.*, 2014). This does not mean that escorting a child to school every day and locking them in class will always improve GPA. Other issues such as family life, history, personal factors and previous success potentially affect both GPA and attendance (Allensworth *et al.*, 2014). Deeper understandings of correlations can lead to the investigation of what is causing the relationship.

Causal analysis focuses on the interplay of two events where the first event is responsible for the second. To invest wisely in improving education, interventions should be based on knowledge of causal relationships where possible. Many unintended consequences arise when acting on only correlational data. Borman *et al.* (2005) point to several educational programs and studies that did not result in the intended outcomes. Another example of unintended outcomes was reported in the popular book *Freakonomics* (Levitt and Dubner, 2006). Daycare centers in Israel depended on the good intentions of the parents to pick up their children on time. When the daycare introduced a financial incentive (i.e. fining parents who were late), parents began to be late more often because they no longer felt guilty for taking the caregiver's time. Double blind randomized controlled trials are the standard for causal analysis. Predictive data modeling can in some cases also provide insight into causal relationships.

When an intervention's outcomes are likely high-cost or high-risk, causal analyses should be the basis of action. Nevertheless, when potential outcomes are low cost and risk, it may not be worth the time and resources required to obtain a causal analysis. To gather data for a causal analysis, randomized assignment of students, groups, classes or schools to a control and an intervention must occur before the data are gathered. Many federal grants (e.g. Gaining Early Awareness and Readiness for Undergraduate Programs and First in the World) provide opportunity for causal analyses. For example, classes may be randomly assigned to receive programs such as financial advising for college. If assignment was random and a control group exists, an appropriate analysis can prove that the program, as administered, did (or did not) affect post-secondary attendance.

Appropriate conclusions based on an analysis
An improper match between the question and the type of analysis leads to sub-optimal outcomes. The curve in Figure 1 represents the optimal use of available analyses. Over-application will lead to risks of unintended consequences and unrealized expectations. Under-application will lead to forgone opportunities to improve education. The literature clearly shows that decisions based on non-empirical data are more likely to include bias and lead to unintended consequences. Because of the current regulatory environment around education in the USA, most educators make large decisions with the support of evaluations, rankings and other basic summary analyses. The implication of this principle of matching the analysis to the question is that there is more potential that can be unlocked from our data stores beyond what naturally happens due to regulation.

Extracting optimal value from the data is done by moving up the curve in Figure 1. Over-application of an analysis can lead to overconfidence in decisions. Correctly applying the data and analysis can lead to the best decisions, given current data resources.

Data competency maturity model
The practice of developing and implementing maturity models has led to processes improvements in small and large organizations and from a wide variety of sector. Organizations that have adopted maturity models for process improvement have seen improvements in cost, schedule (predictability and time required to do tasks), productivity, quality, customer satisfaction and return on investment (benefit-to-cost ratios, net present value, internal rate of return, payback periods and breakeven points). Goldenson and Gibson (2003) report on successes in software development from the application of competency
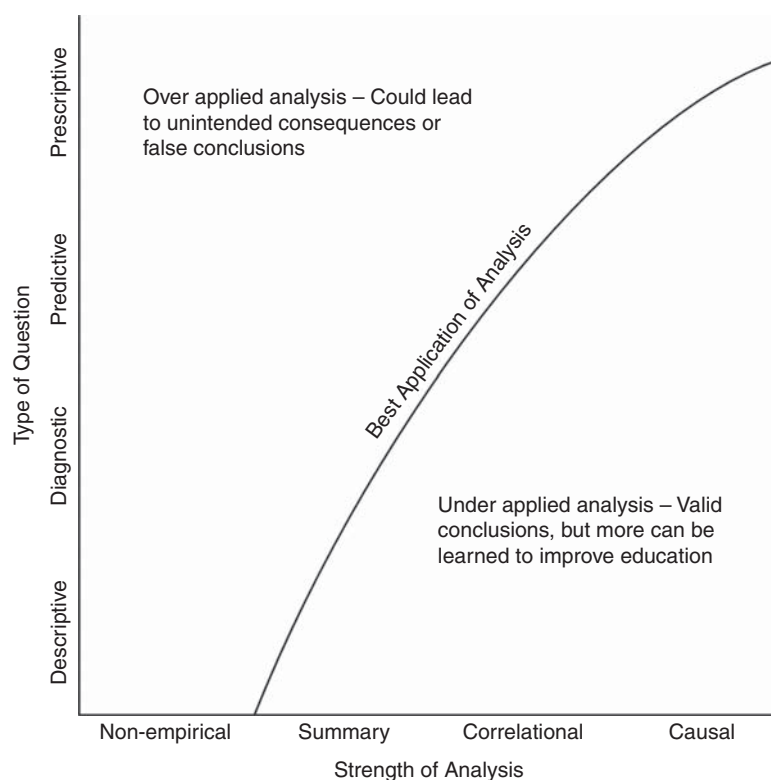
Figure 1.
Matching analysis to
a type of question

maturity models. Boeing reduced release times by 50 percent. General Motors was able to increase the number of development milestones met from 50 to 95 percent. Lockheed Martin increased software productivity by 30 percent. These organizations have achieved improvements in many dimensions (Gibson *et al.*, 2006).

Most maturity models are patterned after a maturity model developed by the US Department of Defense Software Engineering Institute (Humphrey, 1989). Examples include those used in project management (Project Management Institute, 2013) and process management (De Bruin and Rosemann, 2005). Several models have been developed around analytics. The organization Institute for Operations Research and Management Science publishes an analytics maturity model as a self-assessment tool. Other organizations such as SAS, Booz Allen Hamilton, Accenture, Adobe, The Data Warehousing Institute and the Health Information Management Systems Society all have versions of an analytics maturity model. As useful as these various models are in their context, they do not directly address the strength of analysis, the type of data used, and the resources required to leverage that data. The use and application of these models is illustrative of the value of data. The data competency maturity model presented here does not exist in any other form; however, it is rooted in established principles from the academic literature and has the potential to be more robust than those based on industry experience alone. Additionally, there is no model that we know of that provides a roadmap for developing organizational competence in data analytics.

Each stage represents an increase in maturity and capability of a school to leverage data (see Table I). Educators' specific roles and responsibilities will shape their perceptions of

and need for evidence (Coburn and Talbert, 2006). Where data and analytic skills and resources are not available or costs and risks of potential outcomes are low, schools will find lower stages more optimal. Further, the complexity of a project may determine which maturity level is the most optimal (Albrecht and Spang, 2014). Nevertheless, data, tools and analytics skills are becoming more common. Over time, educators should expect and seek to move to higher stages as circumstances change.

The *ad hoc* stage of the data competency maturity model is a state in which decisions are made largely without data. It is likely that data are available in this stage and may be used from time to time. However, data sources are largely unknown to most educators in the organization because data sources are not tracked or managed. Data are disjointed. Analyses of large amounts of data are extremely difficult and time-intensive. There is little attention to or understanding of appropriate application.

The defined stage requires that data sources are identified and cataloged. This allows for regular use of the data. The type of questions that can be answered fully will mostly be descriptive because data are not well integrated at this point. Correlational and causal analyses generally require data from multiple sources. Data users should be aware of the strength of their analysis and apply it appropriately. They must recognize limitations and risks when it is necessary to answer questions above the ability of the analysis to answer. The preparation for the next stage would require that data sources are not just listed, but managed. An effort is made to make sure that data are gathered and stored in a way that the data are clean and useful. Where necessary, data sources are digitized.

The integrated stage requires that data be integrated. Integration can happen through practices of data warehousing. At this point, tools for visualization and summary analyses become available for decision makers who may not have skills in data management. Correlational analyses become possible. However, experience shows that many of the tools that become available to end users only provide summary analyses and visualizations. Therefore, correlational analyses are not common. Developing a culture of data use and evidence-based decision making is now possible because the data can be made directly available to educators and other decision makers.

The optimized stage requires that an organization have access to statistical and data management skill sets. Up to this point, data management has been necessary and is often available to information technology staff. However, the synergy between these skills is critical. These skills could come from a variety of different sources. Some educators in math and statistics may have the ability to fill this need. Using internal talent would often require additional coursework in data management. Larger schools and LEAs may have the justification and ability to hire a data scientist. In some cases, this function could be outsourced for intermittent needs. The person filling this role will have to be aware of discussions among teachers, administrators and of the issues they face. The data scientist

Stage	Description
ad hoc	Minimal data use. Data are disjoint
Defined	Data sources are cataloged and described. Decision makers understand the weakness and strengths of the data they use. Summary analyses are common
Integrated	Data are integrated through tools such as data warehouses and data visualization tools. Correlational analyses are possible. Culture of evidence-based decision making develops
Optimized	Data management and statistical skills are present in the organization. Efforts are made to optimize the use of available data. Diagnostic and predictive applications of correlational data are common
Advanced	Experimentation provides causal data. Institutional review boards become necessary

Table I.
Stages of the data competency maturity model

actively produces analyses to address these issues. The organization begins to optimize the use of all available data. At this stage, the use of data should be built into the routines of the organization. Working the use of data into routines is one potential way to increase the use of data (Little, 2012; Spillane, 2012).

The advanced stage allows for prescriptive questions to be answered supported by causal analyses. Causal analyses require randomly assigned educational or policy interventions. This random assignment of interventions can often have ethical implications in secondary education. Therefore, this stage requires an institutional review board and policies regarding what types of situations are inappropriate for experimentation. The most likely source for these resources is currently partnerships with universities. Most commonly this happens through grants. No Child Left Behind (NCLB) and other legislations have attempted to promote such experimental or quasi-experimental studies (Coburn and Talbert, 2006). Traditional academic research in secondary education suggests that randomization here is a problem. Policy makers, parents and other interested parties often push to have at-risk or advanced students participate in the experiment because they believe there is some benefit in the treatment. Until the experiment is complete, this belief is unfounded. Traditional threats to validity must be addressed at this stage. Beyond general- or system-level policies, personalized learning programs may also be in this stage (e.g. Altschool, 2016).

Table I defines the various stages of the data maturity model. Figure 2 suggests the comparative potential value of each stage due to stronger analyses and leveraging data. The literature suggests that better matching of question to analysis will allow LEAs and
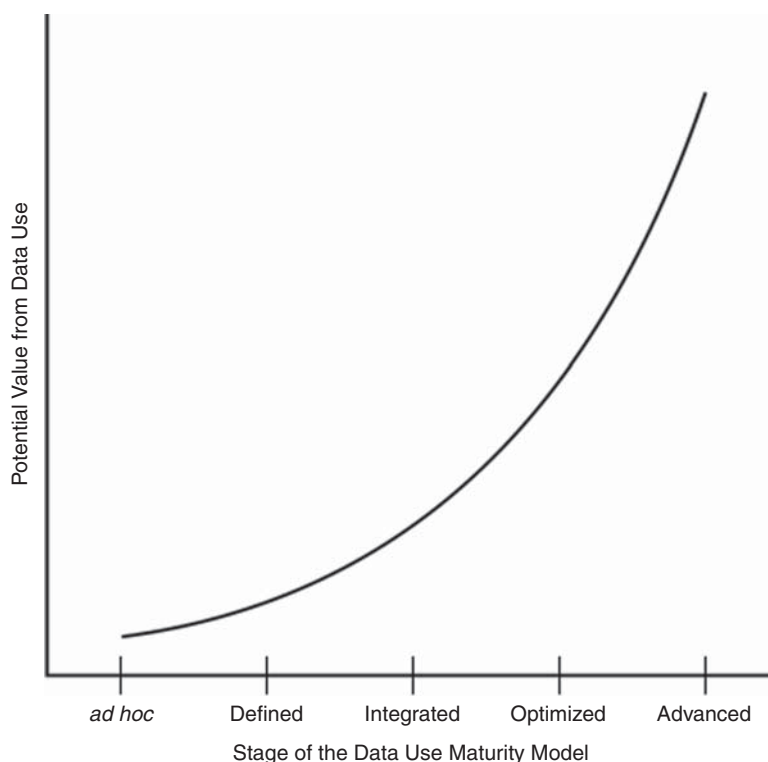

Figure 2.
Suggested
potential value of
operating at a given
maturity model level

schools to extract more value from their data. We assert that development of the skills, culture and processes defined in the data competency maturity model enables educators access to the appropriate analysis. The value of these analytics is expected to increase as illustrated in Figure 2; however, the shape of this return will likely vary by the target performance metric, environment and initial performance of individual organizations. Further work must be done to investigate the additional potential of each step of the maturity model. These returns should give organizations a powerful incentive to increase their efficiency and effectiveness.

Need for more appropriate application of analyses

Higher levels of the data competency maturity model could lead to more confident decision making and stronger outcomes from intervention efforts. Taylor *et al.* (2015) point out that "[m]any stakeholders in education, practitioners in particular, need evidence from rigorous trials about which comprehensive programs […] have the greatest effects on student outcomes" (p. 985). These interventions and programs are often controversial and would benefit from more rigorous application of evidence. Examples could include homework policies, desegregation and standardized testing. More focused and appropriate application of data is necessary to determine which interventions are most effective.

Non-empirical conclusions in homework policies have developed a culture where more homework has generally been associated with more rigor. A new approach has been introduced which focuses on the well-being of the student and limits the quantity of homework and amount of time students spend on it. Recent data suggest that there is no correlation in the amount of homework given and the academic success and well-being of the student (Forestieri, 2015). Homework policies have typically been based on observational data analysis, meaning they overlook the relationships between student homework load, student well-being and student outcomes (Pope, 2010).

In some cases, the use of summary analyses to desegregate schools has not been successful. School choice within San Francisco is driving a re-segregation of schools. More affluent, educated parents compete for the small number of seats at the highest-performing schools. Others end up in under-performing schools. The mechanisms that promote diversity (giving preference to students who live in neighborhoods with below average test scores) have shown fundamental flaws (Smith, 2015).

Correlational analyses traditionally show that higher test scores mean higher success. However, driving up test scores has not always lead to the intended outcome. Studies have explored the policies that financially penalize struggling schools for poor standardized test scores (Welton and Williams, 2015). Many conclude that these policies hinder, rather than improve, a student's college readiness (Welton and Williams, 2015). Policies such as NCLB were intended to make schools more accountable; instead, it has led to a focus on test scores. This stems from a misunderstanding of causality and what is driving student preparedness and how to make schools accountable. This policy failed to consider the correlational relationships between struggling schools, their funding and overall student outcomes (Welton and Williams, 2015). The unintended consequences of NCLB lead to the 2015 replacement, ESSA (2015).

In an ideal world, a high school would have processes in place to capture all relevant data and put it to use. Administrators from this high school would receive regular reports about the status of a variety of key performance indicators (summary analysis). After reviewing the descriptive analysis, the administrators notice the dropout rate is increasing and they would like to identify at-risk students to proactively intervene to reduce the school's dropout rate and increase student success. A more detailed correlative analysis would then take into account student background, academic performance and all other available factors to

predict which students are more likely to drop out. In addition to a list of the students which are most likely at-risk, the model will provide information that may be useful for interventions. A large portion of the at-risk students may be from a certain area and income. Such a finding could point to economic distress, an increase in illicit drug use or violence, or a number of other factors. Such knowledge would lead to potential interventions. Assuming the issue is economic distress in a certain area of the city, programs of peer tutoring, home visits and need assessments, or many other possible programs could be proposed. Educators would select an intervention and implement it. The school would then track who was involved with the intervention and to what level. The program would be assessed on the dropout rate of those students that were most at risk (causal analysis). At this point, if the program was successful, the data will show that implementation of the intervention caused a decrease in dropout rates in this context. The program could then be used confidently and applied to similar contexts.

The current state of data in education

A review of extant literature suggests that many critical and valuable data are available to educators and decision makers. This work does not attempt to assemble an exhaustive list of these data; however, a brief overview provides insight into possible progression through the stages of the data competence maturity model (DCMM). A review of the literature also provides a measure of the application of data and analyses to the issues that educators face.

A core set of factors contributes to the ability of decision makers to improve retention, graduation and post-secondary enrollment. As found in the literature, these factors can be rolled into, first, academic performance data and, second, student demographic and historical data. Other groupings are feasible, but these suffice for evaluation of data availability and use by educators. See Tables II and III for more detail and example references.

Demographic and historical data include socioeconomic status, family size and structure, parent characteristics, social engagement and educational attitudes. Background factors are generally static (e.g. gender or socioeconomic status). Although a substantial amount of demographic and historic data is available, it can be difficult to leverage. For example, data regarding an individual student's free and reduced lunch status are closely guarded, even within schools.

Academic performance data should be leveraged at both the student and organizational levels. A student's academic performance is often related to their retention and graduation (Lotkowski *et al.*, 2004). Analyses on a higher level should also leverage data on school performance because the school has a substantial impact on individual student retention and graduation (Lotkowski *et al.*, 2004).

Further, we can categorize both academic and demographic data by how easy they are to influence or manipulate. To reach the highest levels of data use maturity, policy and program experiments must occur. Some factors are difficult or impossible for educators to change or influence. Other factors are not ethical to manipulate. There is a continuum between factors that are easy to influence and factors that are impossible to influence. Those factors that are more easily influenced include many academic performance metrics such as attendance and frequency of behavior incidents. Unfortunately, many data requirements related to local and federal reporting are factors that are more difficult to influence (e.g. socioeconomic status, grade point average, standardized testing). As a result, the factors that could be influenced more easily are not as readily available to drive decision making throughout the educational system (e.g. extracurricular participation, class schedule, facility resources). Although specific circumstances will vary, Tables II and III provide high-level categorization of these data between those that are possible to manipulate (Table III) and those that are not (Table II).

Retention factors	Sub-factors	Overall effects
Socioeconomic status	Free and reduced lunches[a]	Students with free or reduced lunches were less likely to succeed in secondary education
	Family income[a,b,c]	Students whose families had low income tended to succeed less often than those with high incomes
	Home resources[c,d]	Students with access to many home resources such as books, computers, internet, etc., […] tended to be more successful in secondary education
	Region[b]	Those from a more affluent and urban region tended to do better in secondary education as opposed to those from more rural areas
Family size and structure	Family structure[a,b,c,e,f]	Students from two parent households showed better results throughout their secondary educations
	Teen parent status[a,b]	If the student was a parent then their secondary education suffered
	Number of siblings[b,e]	The more siblings a student has the higher the chance of the student struggling in their secondary education
	Special education or disability status[a]	Those students with special education needs or disability status struggled far more than those without
Parent characteristics	Parent's education[a,b,c]	The level of education that a student's parent achieved was heavily related to the educational attainment of the student
	Immigrant or English learner status[a,f]	If a parent was an immigrant or an English learner the student tended to struggle much more throughout their secondary education
	Parents employment[b,c]	Students whose mothers worked tended to struggle in their secondary education
	Age of mother at birth[e]	Students whose mothers were young when they gave birth tended to do poorly during their secondary education
Educational attitudes	Parent attitudes[d,e,f,g,h,i]	Students whose families provided little to no parental support, supervision, or expectations did poorly in their secondary education
	Student attitudes[c,d,e,g,i,j]	Students with low self-esteem and self-efficacy tended to do poorly in secondary education
Academic performance	Standardized test scores[a,c,e,g,k,l]	Low scores are correlated with poor success in secondary school
	Grade point average[a,d,e,g,h,m,n,o]	Low grade point averages were correlated with poor performance in secondary education
	School mobility[c,f,g]	The more a student moved the less likely they were to succeed in secondary education

Notes: [a]Institute of Education Sciences (2014); [b]Ekstrom *et al.* (1986); [c]Finn and Rock (1997); [d]Roderick and Camburn (1999); [e]Alexander *et al.* (1997); [f]Rumberger (1995); [g]Rumberger and Larson (1998); [h]Barrington and Hendricks (1989); [i]Battin-Pearson *et al.* (2000); [j]Vallerand *et al.* (1997); [k]Linderman and Baron-Donovan (2006); [l]Jimerson (1999); [m]Allensworth and Easton (2005); [n]Allensworth and Easton (2007); [o]Suhyun *et al.* (2007)

Future data use, ethics and controversies
A concrete and complete assessment of where schools and LEAs currently function in the maturity model must come from future research and survey work. However, some insights can be gained into the current state of analytics in education by reviewing regulation, research and experience.

There is no question that data are flooding school systems and LEAs (Horn *et al.*, 2015). NCLB and subsequent legislation (i.e. Every Student Succeeds Act) has required that assessment data, accountability data and teacher quality data be tracked. Therefore, most schools are required to track, store and manage these data. Actual use of the data does not always match the intentions of policy makers (Honig and Venkateswaran, 2012). Even appropriate compliance with regulation does not necessarily move an organization out of the *ad hoc* stage to the defined stage. Many other sources of data exist in schools that are not cataloged or managed. Tables II and III summarize data, some of which are

Retention factors	Sub-factors	Overall effects
Social engagement	Burn out[a]	Students with high educational expectations but few resources will tend to do poorly throughout their secondary education
	Working[b]	Students who work 10 > hours a week tend to do better in secondary education while those who work 10 + per week tend to do poorly
	Extracurricular participation[c,d]	Students who participate in extracurricular activities tend to do better in their secondary education
Academic performance	Rate of school attendance[d,e,f,g,h,i,k]	Students who attend classes more regularly tend to do better
	Frequency of behavior or discipline incidents in school[c,e,h,i,j,k]	Those with fewer disciplinary or behavioral incidents succeed more often in secondary education
	Performance at grade level[e,f,g,l,m]	Students who are behind in their grade level have a difficult time catching up and, as a result, tend to do poorly in secondary education
Academic performance continued	Classes schedule[n]	Students enrolled in more difficult classes tend to do better in their secondary education
	Classes for college credit[o]	Students who are enrolled in concurrent classes or AP classes succeed more in their secondary education
School performance	Guidance feedback[n,p,q]	Students that attend schools with access to guidance and feedback from counselors tend to do better
	Facility resources[a,n,p,q,r]	Students that attend schools with current technology as well as access to counselors and mentors tend to do better
	School characteristics (staff ability levels and school demographics)[c]	Students who attend schools with low teacher/student ratios, crime, and numbers of students tend to do better

Notes: [a]Salmela-Aro and Upadyaya (2014); [b]Marsh and Kleitman (2005); [c]Finn and Rock (1997); [d]Rumberger and Larson (1998); [e](2014); [f]Allensworth and Easton (2007); [g]Barrington and Hendricks (1989); [h]Battin-Pearson et al. (2000); [i]Rumberger and Larson (1998); [j]Suhyun et al. (2007); [k]Jimerson (1999); [l]Alexander et al. (1997); [m]Allensworth and Easton (2005); [n]Norton (2011); [o]Crouse and Allen (2014); [p]Solberg et al. (2007); [q]Hovdhaugen et al. (2013); [r]Kalsbeek and Zucker (2013)

Table III.
Factors easier
to influence

collected and some of which are not. Outside of the specific measures mentioned in the NCLB legislation, data are available in most schools which can show trends in attendance and grades. However, these data are not generally or easily available to practitioners beyond the data that they generate themselves (through assessing their own students). Even when that data are available, it is often difficult for practitioners to use (Horn et al., 2015).

Other data currently being gathered and often not recognized include internet usage and all the associated information. Substantial controversy exists over the use of such data. When schools have unique logins for all students, it would be possible to see how students spend their time online on school computers. Schools who require students to register their mobile devices to use WiFi resources can track the same data for mobile devices. Such internet usage data would not only provide high-level insights into time usage, but in some cases, could provide information on social interactions or dangerous activities. Policies need to be developed around when this type of data use is appropriate.

There is little evidence to suggest that schools are cataloging and managing data in an intentional and thorough manner. Education has been slow to adopt tools such as data warehousing (Wayman and Stringfield, 2006) which helps to integrate data, at least for purposes of analysis. It is evident that many schools deal with systems which are not interoperable (US Department of Education, 2006). This issue plagues industries across the spectrum. Schools which do not have or do not leverage tools to catalog and track data sources must be categorized in the *ad hoc* stage. Those who have started to track their data may find themselves in the defined stage. Many one-time projects through grants or collaborations with a university may appear to be examples of higher levels of maturity.

Nevertheless, if these examples are not built on a foundation of the earlier stages, the changes are not likely to lead to permanent improvements of the use of data across all decision-making opportunities.

A sample of the literature suggests that data are being gathered in association with interventions. However, these data are not standard and are associated with the grants or initiatives that drive them. Table IV provides an overview of a few interventions and a suggestion on how data are being leveraged in conjunction with these initiatives. These initiatives largely produce correlational analyses. Although causal analyses exist, correlational analyses seem to be more common in the literature. This can be due to political, ethical, financial and parental factors which create difficulties in making trials randomized. The explosion of interest in educational data analytics (Denley, 2014) has allowed for many innovations in predicting which students will require help.

As the data become increasingly available, many practical and ethical questions arise. Some recent trends in data collection are viewed as intrusive of privacy. Horn *et al.* (2015) point to the increasing urgency of questioning which data are useful and which are not. It is easy to argue that educators legally have access to data gleaned from assignments and other academic work. It can also be argued that the school should have access to all internet use logs and surveillance data from cameras for the purposes of improving education. However, many students would be uncomfortable with these proposals. Police body cameras are a current example of controversy around needs, practices and policies (Breitenbach, 2015).

How to progress along the data competency maturity model

Though some schools and LEAs leverage their data well, most educators would agree that the data are not being used as well as it could and should be used (Datnow and Hubbard, 2015). The data competency maturity model can guide schools and LEAs on the path to

Retention factors	Intervention treatment	Extent of data use	Type of analysis
Academic performance data	Tutoring and academic enrichment[a,b,c,d]	Extensive	Explanatory
	Summer enrichment programs[a]	Extensive	Predictive
	Computer skills training[a]	None	Non-empirical
	Mentoring[e,f]	Extensive	Explanatory
	Counseling, advising, career planning, and academic planning[a,c,d,g,h]	Extensive	Predictive
	College visits[a,c,f,h]	Extensive	Explanatory
	Materials development and other[a]	Minimal	Summary/non-empirical
	Educational field trips[a]	Minimal	Summary/non-empirical
	Workshops[a,c,d,f,i]	Minimal	Explanatory
	Job site visits and shadowing[a]	Minimal	Explanatory
Educational attitudes data	Student family events[a,h,j]	Minimal	Summary/non-empirical
	Student cultural events[a,h]	Minimal	Summary/non-empirical
	Parent family events[h]	Minimal	Explanatory
	Parent cultural events[h,k]	Minimal	Explanatory
	Parent counseling/advising[h]	None	Non-empirical
	Parent workshops[h,l]	Extensive	Explanatory
School performance data	Enhancing the teacher's qualification[a,j,m,n]	Extensive	Explanatory

Table IV.
Use of data for analytics in the public education system

Notes: [a]Linderman and Baron-Donovan (2006); [b]Forum for Youth Investment (2014); [c]Ward *et al.* (2013); [d]Yampolskaya *et al.* (2006); [e]Monk *et al.* (2014); [f]Mthethwa-Sommers (2013); [g]Camizzi *et al.* (2009); [h]Tillery (2013); [i]Weiher *et al.* (2006); [j]Castillo *et al.* (2010); [k]Gibson and Jefferson (2006); [l]Davis (2000); [m]Murray (2014); [n]Kim (2010)

better leverage their data. Further, it can be used to enhance professional development for practitioners, guide performance improvement practices and develop research agendas. The model suggests essential next steps toward a more optimal use of accumulating data.

The first step to apply any competency maturity model is to assess where an organization currently fits in the model. If schools and LEAs have developed data practices entirely around regulatory rules and requirements, they are likely to be found in the *ad hoc* or defined stages. Organizations in the latter will have cataloged their available data and developed a process to update that catalog.

The second step will be to assess what is required to move to the next stage of maturity and implement that change. Note that it is not generally possible to skip steps in the model. Each stage builds on the seasoned success of the previous stages. Consider the importance of developing a culture of evidence-based decision making and the use of data in the integrated stage. No matter what tools are built or skills acquired by the organization to move to the optimized stage, analytics will not be of substantial use if the culture does not drive the use of that data.

Third, capabilities need to be maintained. Changes in regulation, administration, financial pressures, etc., can create situations in which capabilities developed in earlier stages need to be revisited. Consider the data catalog required in the defined stage. One-time efforts to create the catalog are not sufficient. The computing and data environments change regularly and significantly. The efforts to list data sources, and types of data available in these sources, need to be a recurring process.

The data competency maturity model suggests major milestones in the path to the advanced application of data for each organization. The catalog required in the defined stage will be the first major and intentional step to better leverage data. At the outset, efforts to catalog the data will require both an administrator and a lead IT staff member. Initial data will range widely and include demographic, cognitive and non-cognitive variables, which many schools will continue to have available only on paper. This catalog will help identify which data sources could be digitizing so that the data are in a useful format for analysis. For example, student writing (once exclusively done on paper) is being done more often on the computer. In many cases, schools are using cloud solutions such as Google Docs to share assignments with teachers. Tools and methods are available from multiple vendors that will search student assignments and provide scores regarding academic measures (e.g. grammar and spelling), emotional measures (e.g. self-focus and self-destructive attitudes) and many other factors. Several variables deemed important to understanding student performance are not currently being leveraged and should be explored in future research. Future work should validate the importance of these variables in improving school performance through analytics.

For schools that have already cataloged their data, the next milestone will be to integrate data. At a minimum, steps should be taken to document what must be done to integrate data. Inability to integrate data will, in many cases, stop organizations from executing correlational analyses and thus prohibit strong answers for diagnostic and predictive questions. At this point, educators often find that identification numbers across old and new systems do not match or that additional information is needed to provide strong interoperability. Consider the case of a large group of students where they often share similar or identical names. Information such as the grade, teacher, time of day or birthdays is needed to complete matches between data sources.

Cultural development is a substantial milestone which must happen at the integrated stage. A large body of literature has examined the issues of data use culture and culture development in schools and organizations more generally. This literature is beyond the scope of the current work. However, we suggest that there are successes readily available to help improve the application of data and start to develop a culture of data use

(Anderson *et al.*, 2010; Gannon-Slater *et al.*, 2017). Finding these opportunities to develop a culture of data use requires an evaluation of decision-making practices and the match between question and analysis. This step often produces low hanging fruit where a more appropriate application of data is possible using data that are easily produced. Initial efforts to use data should be leveraged to make data analysis and application easier and build the culture of data usage.

The next milestone will be putting tools in the hands of decision makers or making analytics readily available through a data scientist. Access to these tools also requires time to learn and use the tools. Leveraging performance indicators, trends and other demographics should increase the effectiveness of counsel given to students and their parents. Nevertheless, these educators are often overloaded and time is precious. Proper decision support requires making the right information available to the right person, in the right format, in the right channel, at the right time (Fenton and Biedermann, 2014). Doing this correctly empowers decision makers to make timely, evidence-based decisions. Providing decision makers this information also requires that educators and administrators be data-literate (Mandinach and Grummer, 2013).

Effective application of data will require pushing data analysis to prediction and prescription for individuals. Through a more mature use of data, interventions around many known factors could be better developed. As an example, absenteeism is receiving attention in the press due to a report of high absenteeism rates among high school students (US Department of Education, 2016). Research has shown that absenteeism is related to graduation rates (Messacar and Oreopoulos, 2013). However, applying this knowledge to individuals will require determining risk levels due to specific amounts of days or classes missed.

The production and use of causal data marks a move to the most advanced stage of the model when the skills and capabilities of previous stages are present. Most schools and LEAs will not have the resources to reach this milestone on their own. Some of the most reachable opportunities for the production and use of causal data come through grants obtained in partnership with research universities. The opportunities through GEAR UP and FITW have already been mentioned. Experience shows that even in these situations, great care must be taken to develop processes for appropriate trials of interventions and data gathering. Otherwise, natural forces will cause interventions to be implemented in a manner that only allows correlation to be assessed. For example, it is easy to imagine an intervention where parents are invited to an evening where they are taught about how to manage time and how to help their children manage time to complete their school work. Such an intervention is likely to show that the students whose parents participated did better at completing their work than students whose parents did not attend. However, the parents who self-selected to participate likely have more time, resources and interest in improving their children's ability to complete homework (the parents had the time and made the effort to come to the meeting). Parent's interest is an example of the natural forces which will not allow interventions to be tested without well planned processes in place.

The highest stage is not always the best stage for a given school or LEA. A small, rural school with very little digitized information will not likely find that investment in data science skills is an appropriate use of resources. Additionally, it is not advisable to jump straight to the highest stage. Going through each stage of the process in a diligent way can help build a culture of making data-driven decisions and help LEAs better understand their data. This model provides several benefits in this case as well. First, understanding of where the school sits in the data competency maturity model and what the next step is will help administrators make long-term plans. Second, an understanding of the principles of the appropriate application of data allows educators to better understand when unintended

consequences may appear. The application of the data competency maturity model requires assessment of an organizations current status, identification and implementation of the next steps, and vigilance to assure that capabilities developed in previous stages are maintained. The model itself provides the list of major milestones to move through the stages.

Future research and limitations

The data competency maturity model presented in this paper is still in an initial stage. Future research should focus on the development, implementation, and testing of the model. One of the most difficult challenges to full implementation of data analytics is that many organizations do not have the necessary training and skill sets to use the analyses produced. Can educators be trained to access and use the analyses that come from the data? Can data analysis be done in a way that is not obtrusive to critical planning time? How can the right data and analysis be presented when it is needed? How can all participants be effectively trained in ethical use of the data?

There is still a substantial lack of research on the use of data by practitioners (Coburn and Turner, 2012). We believe that this is due in part to a lack of knowledge of how to develop the capabilities to use data within secondary education organizations. As the model is implemented in secondary education, and possibly other relevant contexts, adaptations of the model will certainly be necessary. Therefore, this model should be subject to discussion and reform as it is used. Outcomes of the model can then be examined as schools and LEAs make progress in this regard.

Further research should also focus on the changes that occur as data analytics is used more in education. Analytics cannot replace educators. To the contrary, as more concrete evidence becomes available to decision makers and as that data are centered closer to the educator, educators should be given more freedom to apply that evidence in their context. Further work should also investigate how decision making shifts as LEAs and schools move along the data competency maturity model.

Debate must continue regarding the appropriate use of data collected in schools. Legislation is an important consideration when dealing with data about students. Currently, the FERPA (1974) restricts data from being used in certain contexts. However, even when data and analyses are kept within the appropriate bounds, there is still the question about student, parent and educator expectations. Ways in which these data will be stored and how it may affect students later in life need to be explored further. The human factor needs to be considered. Legislation will need to create a clear system for how false-positives and false-negatives will affect students. Developing the ability to assess and recognize the potential for unintended consequences needs to be further investigated.

Additionally, improvements in artificial intelligence and data analytics software are making the process of data collection, data cleaning, feature engineering and analysis more accessible to individual users. As this trend continues to evolve, future research should look at how educators can most effectively use these tools to make better decisions inside schools and LEAs. Additionally training and outreach programs can be developed based on these principles and best practices identified though thoughtful application to help school progress through to the appropriate level of the DCMM.

Conclusion

This work proposed a data competency maturity model based on data management, analytic capabilities, a culture of evidence-based decision making and pushing analyses up the continuum to causal evidence. A review of current practices in secondary education reveals that most LEAs and schools have not left the first and second stages. The model provides a roadmap for educators who wish to institutionalize the use

of data to improve educational and organizational outcomes. The model suggests that educators should be given more freedom and autonomy in their environment as they are given better evidence on which to act.

Over the last decade, there has been an explosion of data in secondary education. Using these data correctly can lead to improved outcomes such as graduation rates, post-secondary enrollment, teacher development and teacher satisfaction. Incorrect application can lead to missed opportunities (when data and analyses are under-applied) and unintended and negative outcomes (when data and analyses are over-applied). As technology continues to develop, we must intentionally develop policies and practices to leverage data as a valuable resource for student success. Our data are becoming a valuable and deep resource to improve the lives of students and educators.

References

Albrecht, J.C. and Spang, K. (2014), "Linking the benefits of project management maturity to project complexity: insights from a multiple case study", *International Journal of Managing Projects in Business*, Vol. 7 No. 2, pp. 285-301.

Alexander, K.L., Entwisle, D.R. and Horsey, C.S. (1997), "From first grade forward: early foundations of high school dropout", *American Sociological Association*, Vol. 70 No. 2, pp. 87-107.

Allensworth, E. and Easton, J.Q. (2005), *The On-Track Indicator as a Predictor of High School Graduation*, The University of Chicago Consortium on Chicago School Research, Chicago, IL.

Allensworth, E. and Easton, J.Q. (2007), *What Matters for Staying On-Track and Graduating in Chicago Public Schools*, The University of Chicago Consortium on Chicago School Research, Chicago, IL.

Allensworth, E., Gwynne, J., Moore, P. and Torre, M.D.L. (2014), *Looking Forward to High School and College: Middle Grade Indicators of Readiness in Chicago Public Schools*, The University of Chicago Consortium on Chicago School Research, Chicago, IL.

Altschool (2016), "AltSchool homepage", available at: www.altschool.com (accessed June 17, 2016).

Anderson, S., Leithwood, K. and Strauss, T. (2010), "Leading data use in schools: organizational conditions and practices at the school and district levels", *Leadership and Policy in Schools*, Vol. 9 No. 3, pp. 292-327.

Banerjee, A., Bandyopadhyay, T. and Acharya, P. (2014), "Data analytics: hyped up aspirations or true potential?", *Vikalpa: The Journal for Decision Makers*, Vol. 38 No. 4, pp. 1-11.

Barrington, B.L. and Hendricks, B. (1989), "Differentiating characteristics of high school graduates, dropouts, and nongraduates", *The Journal of Educational Research*, Vol. 82 No. 6, pp. 309-319.

Battin-Pearson, S., Newcomb, M.D., Abbott, R.D., Hill, K.G., Catalano, R.F. and Hawkins, J.D. (2000), "Predictors of early high school dropout: a test of five theories", *Journal of Educational Psychology*, Vol. 92 No. 3, pp. 568-582.

Bienkowski, M., Feng, M. and Means, B. (2012), *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief*, US Department of Education, Office of Educational Technology, Washington, DC.

Borman, G.D., Slavin, R.E., Cheung, A.C.K., Chamberlain, A.M., Madden, N.A. and Chambers, B. (2005), "The national randomized field trial of success for all: second-year outcomes", *American Educational Research Journal*, Vol. 42 No. 4, pp. 673-696.

Breitenbach, S. (2015), *States Grapple with Public Disclosure of Police Body-Camera Footage*, Pew Charitable Trusts.

Camizzi, E., Clark, M.A., Yacco, S. and Goodman, W. (2009), "Becoming 'difference makers': school-university collaboration to create, implement, and evaluate data-driven counseling interventions", *Professional School Counseling*, Vol. 12 No. 6, pp. 471-479.

Castillo, L.G., Conoley, C.W., Cepeda, L.M., Ivy, K.K. and Archuleta, D.J. (2010), "Mexican American adolescents' perceptions of a pro-college culture", *Journal of Hispanic Higher Education*, Vol. 9 No. 1, pp. 61-72.

Cliff, N. (1983), "Some cautions concerning the application of causal modeling methods", *Multivariate Behavioral Research*, Vol. 18 No. 1, pp. 115-126.

Coburn, C.E. and Talbert, J.E. (2006), "Conceptions of evidence use in school districts: mapping the terrain", *American Journal of Education*, Vol. 112 No. 4, pp. 469-495.

Coburn, C.E. and Turner, E.O. (2012), "The practice of data use: an introduction", *American Journal of Education*, Vol. 118 No. 2, pp. 99-111.

Conner, L.J. and Vincent, W.H. (1970), "A framework for developing computerized farm management information", *Canadian Journal of Agricultural Economics*, Vol. 18 No. 1, pp. 70-75.

Cotton, D.R.E., Stokes, A. and Cotton, P.A. (2010), "Using observational methods to research the student experience", *Journal of Geography in Higher Education*, Vol. 34 No. 3, pp. 463-473.

Crouse, J.D. and Allen, J. (2014), "College course grades for dual enrollment students", *Community College Journal of Research and Practice*, Vol. 38 No. 6, pp. 494-511.

Datnow, A. and Hubbard, L. (2015), "Teachers' use of assessment data to inform instruction: lessons from the past and prospects for the future", *Teachers College Record*, Vol. 117 No. 4, pp. 1-26.

Davenport, T.H. and Harris, J.G. (2007), *Competing on Analytics: The New Science of Winning*, Harvard Business School Press, Boston, MA.

Davis, D. (2000), *Supporting Parent, Family, and Community Involvement in Your School*, Northwest Regional Educational Laboratory, Portland, OR.

De Bruin, T. and Rosemann, M. (2005), "Towards a business process management maturity model", *Thirteenth European Conference on Information Systems*, May 26–28, Regensburg.

Denley, T. (2014), "How predictive analytics and choice architecture can improve student success", *Research & Practice in Assessment*, Vol. 9 No. 2, pp. 61-69.

Ekstrom, R.B., Goertz, M.E., Pollack, J.M. and Rock, D.A. (1986), "Who drops out of high school and why? Findings from a national study", *Teachers College Record*, Vol. 87 No. 3, pp. 356-373.

ESSA (2015), Every Student Succeeds Act of 2015, Pub. L. No 114-95 § 114 Stat. 1177 (2015-2016).

Fenton, S.H. and Biedermann, S. (2014), *Introduction to Healthcare Informatics*, AHIMA Press, Chicago, IL.

FERPA (1974), Family Educational Rights and Privacy Act of 1974. 20 U.S.C. § 1232g; 34 CFR Part 99.

Finn, J.D. and Rock, D.A. (1997), "Academic success among students at risk for school failure", *Journal of Applied Psychology*, Vol. 82 No. 2, pp. 221-234.

Forestieri, K. (2015), "High school district to revise homework policy", *Mountain View Voice*, May 15, available at: https://mv-voice.com/news/2015/05/29/high-school-district-to-revise-homework-policy (accessed October 18).

Forum for Youth Investment (2014), *Sharing Data is a Calculated Move: Schools Partner with Afterschool Providers to Improve Youth Services*, Forum for Youth Investment, Nashville, TN.

Gannon-Slater, N., Londe, P.G.L., Crenshaw, H.L., Evans, M.E., Greene, J.C. and Schwandt, T.A. (2017), "Advancing equity in accountability and organizational cultures of data use", *Journal of Educational Administration*, Vol. 55 No. 4, pp. 361-375.

Gibson, D.M. and Jefferson, R.N. (2006), "The effect of perceived parental involvement and the use of growth-fostering relationships on self-concept in adolescents participating in GEAR UP", *Adolescence*, Vol. 41 No. 161, pp. 111-125.

Gibson, D., Goldenson, D. and Kost, K. (2006), "Performance results of CMMI-based process improvement", CMU/SEI-2006-TR-004, Software Engineering Institute, Carnegie Mellon University.

Goldenson, D. and Gibson, D.L. (2003), *Demonstrating the Impact and Benefits of CMMI: An Update and Preliminary Results*, Carnegie Mellon University, Software Engineering Institute.

Grissom, J.A., Rubin, M., Neumerski, C.M., Cannata, M., Drake, T.A., Goldring, E. and Schuermann, P. (2017), "Central office supports for data-driven tallent management decisions", *Educational Researcher*, Vol. 46 No. 1, pp. 21-32.

Honig, M.I. and Venkateswaran, N. (2012), "School–central office relationships in evidence use: understanding evidence use as a systems problem", *American Journal of Education*, Vol. 118 No. 2, pp. 199-222.

Horn, I.S., Kane, B.D. and Wilson, J. (2015), "Making sense of student performance data: data use logics and mathematics teachers' learning opportunities", *American Educational Research Journal*, Vol. 52 No. 2, pp. 208-242.

Hovdhaugen, E., Frølich, N. and Aamodt, P.O. (2013), "Informing institutional management: institutional strategies and student retention", *European Journal of Education*, Vol. 48 No. 1, pp. 165-177.

Hubers, M.D., Schildkamp, K., Poortman, C.L. and Pieters, J.M. (2017), "The quest for sustained data use: developing organizational routines", *Teaching and Teacher Education*, Vol. 67, pp. 509-521.

Humphrey, W.S. (1989), *Managing the Software Process*, Addison-Wesley Longman Publishing Co. Inc., Boston, MA.

Institute of Education Sciences (2014), *WWC Evidence Review Protocol for Dropout Prevention Interventions*, Institute of Education Sciences.

Jimerson, S.R. (1999), "On the failure of failure: examining the association between early grade retention and education and employment outcomes during late adolescence", *Journal of School Psychology*, Vol. 37 No. 3, pp. 243-272.

Jindal, R. and Borah, M.D. (2015), "Predictive analytics in a higher education context", *IT Professional*, Vol. 17 No. 4, pp. 24-33.

Kalsbeek, D.H. and Zucker, B. (2013), "Reframing retention strategy: a focus on profile", *New Directions for Higher Education*, Vol. 2013 No. 161, pp. 15-25.

Kim, T. (2010), "An effective way to improve mathematics achievement in urban schools", *Educational Research Quarterly*, Vol. 34 No. 2, pp. 60-71.

Koch, R. (2015), "From business intelligence to predictive analytics", *Strategic Finance*, Vol. 97 No. 1, pp. 56-57.

Lai, M.K. and Mcnaughton, S. (2016), "The impact of data use professional development on student achievement", *Teaching and Teacher Education*, Vol. 60, November, pp. 434-443.

Levitt, S.D. and Dubner, S.J. (2006), *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*, William Morrow, New York, NY.

Linderman, D. and Baron-Donovan, C. (2006), "Collaborative Programs Research Report: CUNY Gear UP", The City University of New York, New York, NY.

Little, J.W. (2012), "Understanding data use practice among teachers: the contribution of micro-process studies", *American Journal of Education*, Vol. 118 No. 2, pp. 143-166.

Lotkowski, V.A., Robbins, S.B. and Noeth, R.J. (2004), "The role of academic and non-academic factors in improving college retention: ACT policy report", American College Testing (ACT), Inc., Iowa City, IA.

Lowin, A. and Craig, J.R. (1968), "The influence of level of performance on managerial style: an experimental object-lesson in the ambiguity of correlational data", *Organizational Behavior and Human Performance*, Vol. 3 No. 4, pp. 440-458.

Mandinach, E.B. and Grummer, E.S. (2013), "A systematic view of implementing data literacy in educator preparation", *Educational Researcher*, Vol. 42 No. 1, pp. 30-37.

Marsh, H.W. and Kleitman, S. (2005), "Consequences of employment during high school: character building, subversion of academic goals, or a threshold?", *American Educational Research Journal*, Vol. 42 No. 2, pp. 331-369.

Messacar, D. and Oreopoulos, P. (2013), "Staying in school: a proposal for raising high-school graduation rates", *Issues in Science and Technology*, Vol. 29 No. 2, pp. 55-61.

Milliken, A.L. (2014), "Transforming big data into supply chain analytics", *Journal of Business Forecasting*, Vol. 33 No. 4, pp. 23-27.

Monk, M.H., Baustian, M.M., Saari, C.R., Welsh, S., D'elia, C.F., Powers, J.E., Gaston, S. and Francis, P. (2014), "EnvironMentors: mentoring at-risk high school students through university partnerships", *International Journal of Environmental & Science Education*, Vol. 9 No. 4, pp. 385-397.

Mthethwa-Sommers, S. (2013), "Reading beyond research results: the hidden curriculum in a college and urban high schools partnership", *Journal of Community Engagement & Higher Education*, Vol. 5 No. 2, pp. 45-55.

Murray, J. (2014), "Critical issues facing school leaders concerning data – informed decision-making", *Professional Educator*, Vol. 38 No. 1, pp. 1-8.

Norton, M.S. (2011), "Please, not another push to get tough on student retention", *Planning & Changing*, Vol. 42 Nos 3/4, pp. 209-223.

Pope, D. (2010), "Beyond 'doing school': from 'stressed-out' to 'engaged in learning'", *Education Canada*, Vol. 50 No. 1, pp. 4-8.

Project Management Institute (2013), *A Guide to the Project Management Body of Knowledge*, Project Management Institute, Inc., Newton Square, PA.

Reeves, T. (2017), "School level and other differences in Illinois teacher's use of data to inform instruction", *Mid-Western Educational Researcher*, Vol. 29 No. 4, pp. 332-354.

Roderick, M. and Camburn, E. (1999), "Risk and recovery from course failure in the early years of high school", *American Educational Research Journal*, Vol. 36 No. 2, pp. 303-343.

Romanczyk, R.G., Kent, R.N., Diament, C. and O'leary, K.D. (1973), "Measuring the reliability of observational data: a reactive process", *Journal of Applied Behavior Analysis*, Vol. 6 No. 1, pp. 175-184.

Rumberger, R.W. (1995), "Dropping out of middle school: a multilevel analysis of students and schools", *American Educational Research Journal*, Vol. 32 No. 3, pp. 583-625.

Rumberger, R.W. and Larson, K.A. (1998), "Student mobility and the increased risk of high school dropout", *American Journal of Education*, Vol. 107 No. 1, pp. 1-35.

Salmela-Aro, K. and Upadyaya, K. (2014), "School burnout and engagement in the context of demands-resources model", *British Journal of Educational Psychology*, Vol. 84 No. 1, pp. 137-151.

Slavin, R.E., Cheung, A.C.K., Holmes, G., Madden, N.A. and Chamberlain, A.M. (2013), "Effects of data-driven district reform model on state assessment outcomes", *American Educational Research Journal*, Vol. 50 No. 2, pp. 371-396.

Smith, J.A. (2015), *As Parents Get More Choice, S.F. Schools Resegregate*, San Francisco Public Press.

Solberg, V.S.H., Carlstrom, A.H., Howard, K.a.S. and Jones, J.E. (2007), "Classifying at-risk high school youth: the influence of exposure to community violence and protective factors on academic and health outcomes", *The Career Development Quarterly*, Vol. 55 No. 4, pp. 313-327.

Spillane, J.P. (2012), "Data in practice: conceptualizing the data-based decision-making phenomena", *American Journal of Education*, Vol. 118 No. 2, pp. 113-141.

Suhyun, S., Jingyo, S. and Houston, I. (2007), "Predictors of categorical at-risk high school dropouts", *Journal of Counseling & Development*, Vol. 85 No. 2, pp. 196-203.

Taylor, J.A., Getty, S.R., Kowalski, S.M., Wilson, C.D., Carlson, J. and Van Scotter, P. (2015), "An efficacy trial of research-based curriculum materials with curriculum-based professional development", *American Educational Research Journal*, Vol. 52 No. 5, pp. 984-1017.

Tillery, C.Y. (2013), *The Summative Impact of College Access Interventions: A Program Evaluation of GEAR UP North Carolina*, Appalachian State University.

Trochim, W.M.K. (2000), *The Research Methods Knowledge Base*, Atomic Dog Publishing, Cincinnati, OH.

US Department of Education (2006), *Improving Data Quality for Title I Standards, Assessments, and Accountability Reporting*, US Department of Education.

US Department of Education (2016), *New Data Show Chronic Absenteeism is Widespread and Prevalent among All Student Groups*, Department of Education.

Vallerand, R.J., Fortier, M.S. and Guay, F. (1997), "Self-determination and persistence in a real-life setting: toward a motivational model of high school dropout", *Journal of Personality and Social Psychology*, Vol. 72 No. 5, pp. 1161-1176.

Ward, N.L., Strambler, M.J. and Linke, L.H. (2013), "Increasing educational attainment among urban minority youth: a model of university, school, and community partnerships", *The Journal of Negro Education*, Vol. 82 No. 3, pp. 312-325.

Warner, T. (2014), "9 Reasons you should have a data integration plan", available: http://blog.safe.com/2014/04/why-data-integration/ (accessed June 27, 2016).

Wayman, J.C. and Stringfield, S. (2006), "Data use for school improvement: school practices and research perspectives", *American Journal of Education*, Vol. 112 No. 4, pp. 463-468.

Weiher, G.R., Hughes, C., Kaplan, N. and Howard, J.Y. (2006), "Hispanic college attendance and the state of Texas GEAR UP program", *Review of Policy Research*, Vol. 23 No. 5, pp. 1035-1051.

Weinbaum, E.H. (2015), "State test data and school improvement efforts", *Educational Policy*, Vol. 29 No. 3, pp. 478-503.

Welton, A. and Williams, M. (2015), "Accountability strain, college readiness drain: sociopolitical tensions involved in maintaining a college-going culture in a high 'minority', high poverty, Texas high school", *High School Journal*, Vol. 98 No. 2, pp. 181-204.

Wilson, E. and Demers, M. (2014), "Revolutionary and evolutionary approaches to leveraging predictive business analytics", *Journal of Business Forecasting*, Vol. 33 No. 4, pp. 4-10.

Yadav, R. and Kumar, T. (2015), "Usage of big data analytics for customer relationship management", *International Journal of Advanced Research in Computer Science*, Vol. 6 No. 2, pp. 1-3.

Yampolskaya, S., Massey, O.T. and Greenbaum, P.E. (2006), "At-risk high school students in the 'gaining early awareness and readiness' program (GEAR UP): academic and behavioral outcomes", *The Journal of Primary Prevention*, Vol. 27 No. 5, pp. 457-475.

Corresponding author
Trent J. Spaulding can be contacted at: spauldingtj@appstate.edu