IMPROVING INFORMATION PRODUCTS FOR SYSTEM 2 DECISION SUPPORT

A Dissertation Submitted

to the Graduate School

University of Arkansas at Little Rock

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in Applied Science

in the Department of Applied Science

of the Donaghey College of Engineering and Information Technology

May 2010

Neal Gibson

B.A. University of Arkansas, 1988

M.A University of Arkansas, 2004

M.S. University of Arkansas Little Rock, 2007

This dissertation, "Improving Information Products for System2 Decision Support," by Neal Gibson, is approved by:

Dissertation Advisor:

John Talburt
Professor Information Science

Dissertation Committee:

Elizabeth Pierce
Associate Professor Information Science

Mihail Tudoreanu
Associate Professor Information Science

Ningning Wu
Associate Professor Information Science

Xiaowei Xu
Professor Information Science

Graduate Coordinator:

Abhijit Bhattacharyya
Associate Professor Applied Science

Graduate Dean:

Patrick Pellicane
Professor of Construction Management

Fair Use

Duplication

Signature____

IMPROVING INFORMATION PRODUCTS FOR SYSTEM 2 DECISION SUPPORT, by Neal Gibson, May 2010

ABSTRACT

The creation, maintenance, and management of Information Product (IP) systems that are used by organizations for complex decisions represent a unique set of challenges. These challenges are compounded when the purpose of such a systems is also for knowledge creation and dissemination. Information quality research to date has focused mainly upon treating IP independent from the actual users, despite the obvious interdependency between the two. Research in cognitive psychology has established a dual-process model for human cognition. Designing IP systems in recognition of these differing methods of human cognition represents a new approach to improving their quality. Education data and the decisions that need to be made from such data represent a task environment that cognitive psychologists label as "System 2;" multifaceted decisions needing to be made from complex data, with little agreement on the solution set. This research demonstrates the efficacy of designing IP systems specific to System 2 decision support by the creation of a new application specific to education data and evaluating user responses as to its fitness for use.

Acknowledgments

I have been blessed to be around extraordinary people all my life. Specific to this endeavor, I feel compelled to give special thanks:

TO

…Jim, whose wisdom I will always seek.

…JRT and Greg, who travel the same winding path.

…my father Lester, who taught me the power of work and laughter

…my mother Sue, who taught me the power of sacrifice and creation.

…my children, Rainey and Caleb, for which the word "wonderful" was created.

…and my wife Jann, the muse who has my heart forever.

May God bless.

Table of Contents

List of Tables and Figures

Improving Information Products for System 2 Decision Support

Introduction

The Problem Domain

The creation, maintenance, and management of Information Products (IP) systems that are used by organizations for complex decisions represent a unique set of challenges. Consumers of such systems, when faced with what they perceive to be an overwhelming amount of information and decision choices, may resort to cognitive shortcuts with devastating results. Improving such systems requires that developers lessen the cognitive requirements on consumers by focusing on making the information presented readily understandable and providing the context to make it relevant. These challenges are compounded when the purpose of such IP systems is also for organizational knowledge creation and dissemination. Since institutional knowledge is the goal of such products, another important dimension is that tacit individual and organizational knowledge be captured and made explicit. Quality initiatives directed toward improving systems used in complex decision processes for organizations should be focused on easing the cognitive load associated with use of such IP systems, continually monitoring use to ensure proper context is provided, and facilitating the sharing of information among a community of practice.

To test the efficacy of the ideas represented in this research, a new IP system was created to support complex decision processes and to help serve the creation and management of organizational knowledge. Although this system was specific to a particular domain, education data, it is believe that the lessons learned are applicable to a wide range of information needs specific to complex decisions requiring careful rational deliberation—what cognitive psychologists term "System 2 decisions." The IP system was constructed in an incremental and iterative manner, in which user feedback was central to shaping the overall development. A survey instrument was constructed to measure consumer perceptions of this new system and to compare it to similar existing systems. The overwhelming positive feedback from users of the IP system developed in this manner suggests that developers of systems specific to System 2 decision processes and organizational knowledge would be well served to follow a similar approach and that much more research in IQ should be devoted studying consumer's use of such systems.

There is certainly not a problem with too little data for educators to act upon. As will be shown, the federal government has made the analysis of data by educators an important initiative and has funded these initiatives as well. The problem is instead related to access, but not in the sense most often referred to in IQ research. Educators have access to data, but they do not have access to data in a way that most can find readily useful. This is certainly not a problem specific to education. A famous report from 2005, the research firm Gartner predicted that:

> …through 2007, more than 50 percent of data warehouse projects will have limited acceptance, or will be outright failures, as a result of a lack of attention to data quality issues… IT organizations still build data warehouses with little or no business involvement… solving problems the users do not understand. ("Gartner Says More Than 50 Percent of Data Warehouse Projects Will Have Limited Acceptance or Will Be Failures Through 2007", 2005)

This report identifies two key areas of concern for developers; 1) the quality of the inputs used for the creation of the IP systems and 2) adapting the system to user needs. There is a wealth of research in Information Quality (IQ) that address the first, strategies for improving the quality of data inputs and the processes that maintain and transform them, but there is very little research on the second, how to ensure users can effectively use our systems for effective decisions.

IQ has considerable overlap with other academic areas. In the case of data use, there are existing cannons of knowledge in the areas of Operational Research, Cognitive Psychology, and Decision Sciences. Researchers in these disciplines include Nobel laureates such as Herbert Simon and Daniel Kahneman, but virtually no research specific to IQ links the work in these other disciplines to improving IP systems by focusing on the use of data. Research in these other academic disciplines are much more specific to decision making, so it is somewhat natural to look there for insight on how to ensure our products are truly "fit for use," that is, suitable for consumers to use for decisions. There is some research in IQ as to the effect improving data inputs has on decision making, but Ge and Helfert's review of IQ research points out that "interaction and information presentation, which are two important factors influencing decision making, need to be investigated as independent variables in the research of IQ effects on decision making." (Ge & Helfert, 2007, p. 13)

As an academic discipline, IQ borrows heavily from the application of quality initiatives in manufacturing pioneered by Walter Shewhart and Edwards Deming. However, application of their "plan-do-study-act" (PDSA) cycle to IQ is not as natural a fit as might first be assumed, because again, the focus is on data inputs and not the use of the data itself. For example, *Introduction to Information Quality* has a chapter devoted to the standard tool used in manufacturing for PDSA, controlled charts, but the application of sampling to databases seems somewhat misplaced, since often times it would be much faster and easier to simply access the entire collection of a particular attribute and apply measurement to that than to take the time to instead develop a process for representative sampling. Certainly a much more appropriate application for measuring the quality of attributes within a database would be to do the opposite of sampling, for example using resampling techniques such as bootstrapping to infer the impact of poor data quality for such attributes over time.

Every professional that works with data on a regular basis can speak to the problems associated with poor data inputs, and this obviously plays a critical role in improving the quality of such IP systems. However, this concentration on data inputs tends to make IQ as an academic discipline and industry methodology similar in practice to those associated with IS0 9000, which are process-oriented instead of performance-oriented. Such an approach led Henkoff to comment that “With ISO 9000, a manufacturer of concrete life jackets could be registered as long as there were systems in place to assure they were well made.”(Henkoff, 1993, p. 117) IQ research and practices may help us put in place standards and process to improve the creation and maintenance of IP systems, but it does not provide us with metrics associated with the products themselves specific to their use. As with ISO 9000, Rufe points out that “it is important to emphasize that having a certified quality system in place does not guarantee the quality of the system’s outcome.”(Rufe, 2002, p. 317)

It would be difficult to imagine any quality initiative in manufacturing that did not place an equivalent weight on measuring the quality of the final product and feedback from consumers as it does on the components used in its production.   However, in IQ there is little research on how to ensure that the use of any particular IP system leads to good decisions and also little research on how best to illicit feedback from consumers about their use of the system. Redman agrees that fitness of use extends to "their intended use in operations, decision making, and planning." (Redman, 2001, p. 71) It is also important that use of our IP system is an important driver of quality for the inputs themselves. As Orr points out, data quality is a "function of its use, not its collection," and that data quality "will ultimately, be no better than its most stringent use."(Orr, 1998, p. 68)

How then do we best extend quality initiatives for IP systems to include measures specific to its use by consumers, and more importantly, its fitness of use in decision making? We can look to research specific to decision making in other academic disciplines to help us understand the problems associated with decision making and construct measures of conformance to those norms. Specifically, we can look to research in decision making and cognitive psychology that clearly shows that when confronted with either a decision that involves complex data or a complex decision environment, decision makers will often rely upon shortcuts for such decisions, gut instinct, which can sometimes have terrible consequences.

A final consideration for ensuring fitness of use is that many times our IP system will also be needed by an organization for knowledge creation and dissemination. In the only book devoted to both IQ and organization knowledge management, Huang, Lee, & Wang recognize the importance for organization to have a process for what they term "'knowledge hunting…', collecting knowledge, harvesting the process of filtering, and hardening the process of structuring tacit, useful knowledge into explicit, reusable knowledge." (Huang et al., 1999, p. 114) This presents yet another problem in the sense that such systems will have to be constantly evolving, in that as organizational knowledge changes, the data and information systems of the organization will also need to evolve in support of changing understandings about the world.

To build an IP system adaptive to ever-changing views of the world requires the adoption of iterative and incremental development (IID) methodology prevalent in modern software development frameworks. Each iteration of development builds upon what was learned in previous increments, and this learning comes from both the development and use of the system. Since business value, that is functionality, is part of each iteration, IID avoids the problem mentioned in the Gartner report, where systems are built for solving problems users do not understand. Such discontinuity is often the result of a development process where the business value and functionality is delivered all at once, at the end of the development process, at which time changes to the system are very difficult to implement. In contrast, IID can also be seen as an application of PDSA, creating the IP system as a form of evolutionary advancement, continual improvement:

> IID grew from the 1930s work of Walter Shewhart, a quality expert at Bell Labs who proposed a series of short "plan-do-study-act" (PDSA) cycles for quality improvement…. Tom Gilb and Richard Zultner also explored PDSA application to software development in later works. (Larman & Basili, 2003, p. 47)

Chapter 2 outlines gaps in current IQ research, specifically a lack of focus on the actual use of systems by users. It goes on to explain what cognitive psychologists call a "dual-process model for human cognition," and why developers of IP systems should be aware of the different ways humans make decisions. The concepts of bounded rationality and fundamental computational bias will be demonstrated in regards to this dual-process model. Finally, research on the social nature of knowledge will be discussed as another important consideration in the development of IP systems. Chapter 3 will frame the task environment for education data in terms of the IQ gaps outlined above, and a measurement for student growth will be introduced as a necessary data element missing in many education data systems. Chapter 4 discusses the creation of a new IP system for education data in Arkansas, and how its incremental and iterative development was driven by consumer use. Chapter 5 details the results of a survey instrument to evaluate users' impressions of this new application relative to existing systems. And finally, Chapter 6 discusses the implications of this research, including its relevance to other domains and areas for continued research.

Chapter 2

Human Factors and Information Quality

Gaps in Current Information Quality Research

The academic discipline of Information Quality (IQ) has generated a wealth of research on various topics, but little research is specific to studying how data is actually used by clients and how IP systems might be constructed to facilitate better use. Klein, Goodhue, and Davis note that users of information systems are very poor at detecting IQ problems and that researchers “need to develop better theories of human error detection and to improve their understanding of the conditions for improving performance.”(Klein, Goodhue, & Davis, 1997, p. 169)   Ge and Helfert’s review of IQ research points out that a major focus of IQ research is in the effects of IQ on decision making, but that “interaction and information presentation, which are two important factors influencing decision making, need to be investigated as independent variables in the research of IQ effects on decision making.” (Ge & Helfert, 2007, p. 13)

While research specific to IQ is not necessarily new, its immaturity as an academic discipline is perhaps best evidenced by the inability to consistently define terms, specifically definitions for and the relationships between “data,” “information,” and “knowledge.” *Introduction to Information Quality* asserts that “Since there are many levels and interpretations of the differences between data and information, we will treat data and information interchangeable. The context will make it clear.” (Fisher, Lauria, Chengalur-Smith, & Wang, 2006, p. 3) In contrast, English insists that to define IQ one must first clearly delineate between data and information. He writes that data is “the representation of facts about things,” and that information is “data in context.” (English, 1999, p. 18-19) Redman notes the conflicting definitions for both data and information. For data, he finds it more satisfying to indentify data as a datum collection, and each datum itself is a triple of entity, attribute, and value. (Redman, 1992, p. 20) For Redman, defining information is even more problematic. To get around this problem, he identifies a concept of "signals" as being more primitive than data, and that within a collection of signals, "the nonredundant parts are by definition 'informative' and the content is 'information.'" (Redman, 1992, p. 37)

In *Quality Information and Knowledge,* Huang, Lee, & Wang announce that "the words data, information, and knowledge are used somewhat interchangeably in this book," even though earlier in that same paragraph they admitted that "we all agree that the transformation of data for clearer and more meaningful information to users is important." (Huang et al., 1999, p. 146) It is difficult to support on one hand that the terms "data," "information," and "knowledge" are interchangeable while on the other acknowledging that there is a transformation process which turns data into "meaningful information." Despite their discounting a clear delineation of terms between data and information, this transformation process would have to be an important design consideration in the creation of IP systems that support the domain of knowledge creation. This then presents us with another criterion when we design systems for supporting organization knowledge, in that we must present data in a way that our consumers can more readily transform it into information.

Other authors writing about IQ and knowledge are not so dismissive of the differentiation between data and information. Writing from a sociological perspective, Diemers also identifies the transformation of data into information, writing that data is "sensitorily perceptive phenomena in our social world," that becomes information when "data become meaningful and thus relevant to an individual." (Diemers, 1999) In a similar fashion, Davenport and Prusak describe data as "a set of discrete, objective facts about events." (Davenport & Prusak, 1998, p. 2) They also go on to explicitly define the transformation process through which it becomes information:

> Unlike data, information has meaning… Not only does it potentially shape the receiver, it has a shape: it is organized to some purpose. Data becomes information when its creator adds meaning. We transform data into information by adding value in various ways. (Davenport & Prusak, 1998, p. 4)

Davenport and Prusak go on to describe another transformation, where information is transformed into knowledge. It is important to realize that both of these transformations, from data to information and from information to knowledge, are both internal to the individual, and as such, we have no direct measures. Diemers believes that

> individual knowledge is structured in a system of relevancies and typicalites…. The transformational process is then the process of acquiring knowledge out of information which I'm experiencing in my daily life-wrold. (Diemers, 1999)

Davenport and Prusak make a much more concrete link between information and knowledge:

> Knowledge derives from information as information derives from data. If information is to become knowledge, humans must do virtually all the work. This transformation happens through *C* words as:
> *Comparison*: how does this information about this situation compare to other situations we have known?
> *Consequences*: what implications does the information have for decisions and actions?
> *Connections*: how does this bit of knowledge relate to others?
> *Conversations*: what do other people think about this information?

If we accept these explanations as to the links between data, information, and knowledge when creating our IP system, we now have three separate requirements if we wish to address the overall quality of a system used to support organizational knowledge. We must first ensure that the underlying data conform to whatever dimensions of quality we deem important, such as accuracy, timeliness, completeness, etc. We must also contend with the fact that these data will go through a transformation process to become first information and then finally knowledge, so we must facilitate these transformations. However, the most vexing problem confronting the developers of such IP systems is that these transformation processes are internal to the individual for which we can have no direct measurement.

It would seem apparent that in IQ research the actual use of information products would be a central focus. Instead, much of IQ research is devoted to developing universal postulates for IQ which applies to all information products, regardless of their application. This may represent good scholarly research, but it is less helpful for those needing to develop applications and continually improve their quality. In 1972, long before the development of ubiquitous computation, Ivanov laments the fact that while there was a great deal written about IQ, there was still little in the way of practical applications:

> After a review of the EDP [electronic data-processing] literature we find ourselves in a really bad shape. Nowhere is told us how to measure quality and for what purposes, in an explicit manner. We are not able to use the implicit definitions in their present form as a basis for binding negations between a “buyer” and a “seller” of information. To the extent that the authors offer recommendations on what should be done in order to improve quality, we do not know why we should place confidence in their advice; and even if we placed confidence and implemented their advice we would not be able to evaluate the results of their recommendations.(Ivanov, 1972, p. 19)

Almost forty years later, Ge and Helfert's review of the literature concerning IQ identify four central questions that have yet to be answered in the context of applying IQ to actual products:

> What is the relationship between IQ and application contexts?
> How does IQ impact applications contexts?
> What is the relationship between IQ research and information systems research?
> How to control extraneous variables in IQ experiment? (Ge & Helfert, 2007, p. 13)

Despite all the research, journal articles, and books dedicated to IQ, employing the processes of quality improvement to actual IP applications and the use of these applications is somewhat lacking.

In an influential study on the aspects of IQ that are important to consumers, Wang and Strong also note the difficulty in assessing quality of the application and that at least some dimensions of IQ related to products are a function of their use:

> … contextual DQ [data quality] was not explicitly recognized in the data quality literature… Our grouping of dimensions for contextual DQ revealed that data quality must be considered with the context of the task at hand. This was consistent with the literature on graphical data representation, which concluded that the quality of a graphical representation must be assessed with the context of the data consumer's task.
>
> Since tasks and their contexts vary across time and data consumers, attaining high contextual data quality is a research challenge. (Wang & Strong, 1996, p. 20)

Wang and Strong are much more confident in identifying attributes related to "intrinsic data quality," but in a paper directed at Wang and Strong, Gackowski argues that there are no intrinsic qualities to data, they are all contextual:

> By the law of relativity of operations quality, defined as "fit for use," *non-contextual* aspects of quality do not exist. The task-specific required levels

> of accuracy or objectivity are contextual…. The idea of quality attributes intrinsic to data values should be abandoned. (Gackowski, 2006, p. 111)

Gackowski also asserts that "quality requires a rigorous distinction between data and information values."(Gackowski, 2006, 100) However, after defining the difference between the two, he does not specify any particular examples where these differences might come into play and instead combines the two into a single term of "data/information" or simply "D/I" for the rest of the paper.The one thing that does bind these two contrasting views together is that they are both ontological in nature and do little to help inform us as to how we can improve the use of our IP system.

Ontological arguments that are not specific to the type of application in question give rise to this lack of consensuses concerning definitions. Even though Wang & Strong and Gackowski contradict each other concerning quality dimensions that are intrinsic to data, they do recognize that differences in context arise from the nature of the tasks involved. If instead we do take into account that there are differing types of IP systems specific to the types of tasks they support, we can help resolve some of the confusion concerning definitions for "data" and "information." More importantly, if we categorize systems specific to their use, we can look to other academic disciplines for research on how best to shape our products for their intended use.

System 1 and System 2 Decision Processes

The inconsistency of terms in IQ research is a product of ontological requirements to find definitions and processes that apply to the development, maintenance, and management of all type of information products. However, there are multiple types of IP systems and their application. We can classify two broad types that have contrasting characteristics based on the application they are designed to support. To differentiate between these two broad categories, we will borrow from research in social and cognitive psychology where a dual-process model for how humans process information has been developed. Chaiken and Trope summarize the differences between these two models:

> Although these theories differ on a number of dimensions, including domain of application and specific definitions, they all share the basic assumption that two qualitatively different modes of information processing operate in making judgments and decisions and in solving problems. In essence, the common distinction in dual-process models is between a fast, associative information processing mode based on low-effort heuristics, and a slow, rule-based information-processing mode based on high-effort systematic reasoning. (Chaiken & Trope, 1999, p. ix)

Stanovich labels these two differing models for human decision making as “System 1” and “System 2,” with System 1 processes being characterized as “automatic, largely unconscious, and relatively undemanding of computational capacity,” while System 2 is characterized by “controlled processing” or what information theorists term “analytic intelligence.” (Stanovich, 1999, p. 144)

IP systems created for the support of System 1 processes would be products created for automated decisions with a finite class of outcomes. In many cases, the consumers of such products are machines. An example would be an automated payroll system that periodically sends messages to other computers for the purpose of direct-deposit while maintaining an internal record of these transactions for audit and tax purposes. While humans may routinely examine such records, the primary users of such products are machines, and their utilization of these products creates additional data for the creation of yet even more products, such as a bank's own accounting system. This is what gives rise to the saying that "one person's data is another person's information," as well as the difficulty in delineating between data and information. The consumers of products for System 1 processes do not have to be machines, but the decision process they support are automatic in nature with a clearly defined solution set.

An example of a System 1 type IP system for human consumption would be the Check Engine light on your car. The decision that this supports is relatively automatic in nature, in that designers of the car want to give the user a simple indication for what is quite possibly a very serious problem. Despite the simplicity of decisions System 1 products serve, we have all most likely seen numerous examples of poor IQ in such products, for example, a "Check-Engine" light that continues to display due to a bad sensor. The existing body of knowledge about quality improvement adapted from the manufacturing world map easily to System 1 products, since the outcomes of the decisions from these systems can be readily evaluated. We can "close the loop" as it were, between the quality of the IP system and the quality of the decisions stemming from the use of such products.

A far different type of IP system would be those products created for human consumption within problem domains that require complex decisions and where there is often likely no clear correct solution. We will label these types of IP systems as "System 2," after the conscious, multidimensional decision processes humans must employ in such scenarios. An example of such a product would a "buyer's guide" to help people sort through all the possible combinations of price versus features for the purchase of a consumer item. We are confronted by such decisions every day, whether it be choosing what cereal to buy for our children or which outfit we should pick out for work. These types of information products are a much looser fit as we attempt to apply the teachings of Total Quality Management to our decisions, since the outcomes of such decisions are not readily classifiable as correct or not. Even the setting for the decision may be unclear, such as what elements should be considered when making a decision and the general domain of possible solutions. We can certainly apply quality improvement processes to the inputs used for the buyer guide, and we can most likely also get a consumer's impressions of our buyer's guide. The expected utility model can also give us a metric for determining how closely a consumer's final decision fit optimal norms. However, applying these metrics to real-life situations can quickly become problematic. We may decide upon the optimum cereal for our children, only to find out they do not like the one we picked out for them and refuse to eat it. In addition, we will most likely never be able to gather enough data to definitively determine that given all available inputs of price, ingredients, endorsements from athletes or medical associations, etc., that Cheerios are a much better cereal for our children than Wheaties. Worse yet, if our children refused to eat anything but a cereal that all experts in nutrition agree is the worst possible choice for developing bodies, we would be faced with another System 2 decision as to which was worst--for our children to eat a cereal that was bad for them or for them to skip breakfast entirely.

The challenges to human cognition represented by System 2 decision processes are quite familiar in other disciplines. Economics has long recognized the problems associated with trying to understand human rationality, or the lack of rationality, in order to understand economic activity as a function of individual choice. Von Neumann and Morgenstern acknowledged this difficulty as they developed their own model of human rationality:

> The analysis is concerned with some basic problems arising from a study of economic behavior which have been the center of attention of economists for a long time. They have their origin in the attempts to find an exact description of the endeavor of the individual to obtain a maximum of utility… It is well known what considerable—and in fact unsurmounted—difficulties this task involves given even a limited number of typical situations… (Neumann & Morgenstern, 1955, p. 1)

Their expected utility theory goes on to define the four axioms that define a rational decision maker—completeness, transitivity, independence, and continuity. ("Expected utility hypothesis", n.d.) However, expected utility is a normative description of human behavior, and actual human behavior is something quite different:

> The objection could be raised that it is not necessary to go into all these intricate details concerning the measurability of utility, since evidently the common individual, whose behavior one wants to describe, does not measure his utilities exactly but rather conducts his economic activities in a sphere of considerable haziness. (Neumann & Morgenstern, 1955, p. 20)

Indeed, we will demonstrate where human rationality can become quite confused, especially when confronted with an IP system that is unsuitable for the task environment the decision process requires.

The distinction between these two types of IP systems are the decision processes they are created to support, whether they be automated decisions with clearly defined outcomes or decisions requiring tradeoffs between multiple dimensions where the expected outcomes are neither timely nor necessarily well defined. We can imagine that someone playing chess is confronted with numerous examples of System 2 decisions throughout a game, yet we also know that computer systems have been designed that have defeated even world champions, such as the somewhat controversial match in 1997 where IBM's Deep Blue defeated Gary Kasparov. Despite the amazing processing power of Deep Blue and the sophistication of its programming, it still approached every move in a manner that we can model in a series of "if-then" statements, regardless of the number of such statements each move required and the speed by which they were processed. Deep Blue's decision processes, as is the case of all machine-based decision processes currently, are decidedly System 1 in nature.

Although machines must approach all problem domains with System 1 tools, humans will often apply System 1 processes to problems that are clearly System 2 in nature. They do this because System 2 processes represent a much higher cognitive load—we often look upward or close our eyes to focus when thinking deeply about a particular subject. Research suggests that we have an evolutionary aversion to thinking so deeply that we disconnect from the real world:

> Our minds do not seem made to think and introspect; if they were, things would be easier for us today, but then we would not be here today and I would not have been here to talk about it—my counterfactual, introspective, and hard-thinking ancestor would have been eaten by a lion while his nonthinking faster-reacting cousin would have run for cover…. Evidence shows that we do much less thinking than we believe we do—except, of course, when we think about it.(Taleb, 2007, p. xxii)

Simon terms the decisions made from available data as "bounded rationality," that our rational behavior "is shaped by a scissors whose two blades are the structure of the task environment and the computational capabilities of the actor." (Simon, 1990, p. 7) When confronted with an overwhelming task environment, such as those represented in System 2 environments, humans will often take cognitive shortcuts, heuristics, to decrease the cognitive load. Gigerenzer explains how this approach works and why it is an appealing alternative:

> Heuristics are frugal—that is, they ignore part of the information. Unlike statistical optimization procedures, heuristics do not try to optimize (i.e., find the best solution), but rather satisfice (i.e., find a good-enough solution). Calculating the maximum of a function is a form of optimizing; choosing the first option that exceeds an aspiration level is a form of satisficing. (Gigerenzer, 2008, p. 20)

As will be noted latter, the reliance on heuristics can lead to very poor decisions, but it is important to realize that if we are to create IP systems for decision making we must concern ourselves with both the task environment in which they will be used AND the relative cognitive abilities of its intended consumers, including this tendency for users to rely on heuristics when the task environment seems overwhelming.

Dividing the types of IP systems into two broad categories helps us to understand the lack of consensus among IQ researchers on such simple terms as “data” and “information.” More importantly, it also helps us understand why traditional approaches to quality management may be successful to one type of IP system and much less so to the other, simply because decision outcomes are readily apparent and measureable in one but not in the other. While we can imagine a host of differing systems and their application, classifying them into two broad types based on the dual-process model of human cognition helps us map our quality initiatives to the actual decision processes our consumers will use.

The Fundamental Computational Bias

There is a tremendous amount of research that shows how our reliance on heuristics can easily lead us to very poor decisions. Consider the following:

> Jack is looking at Anne, but Anne is looking at George. Jack is married, but George is not. Is a married person looking at an unmarried person?
> A. Yes
> B. No
> C. Cannot be determined

This simple scenario was presented to over three-hundred workshop participants, and well over 99% answered "C. Cannot be determined." (Gibson, 2010) In reality, Anne's marital status is not relevant. If Anne is married, then a married person, Anne, is looking at an unmarried person, George. If Anne is unmarried, then a married person, Jack, is looking at an unmarried person, Anne. Because we rely on heuristics for many computational tasks, we often fail to envision all possible solution sets.

Stanovich, labels such rational failures the "fundamental computational bias," since they are so pervasive and can have such devastating results. He classifies the four main areas of bias as

> 1) the tendency to contextualize a problem with as much prior knowledge as is easily accessible, even when the problem is formal and the only solution is a content-free rule;
> 2) the tendency to "socialize" problems, even in situations where interpersonal cues are few;
> 3) the tendency to see deliberative design and pattern in situations that lack intentional design and pattern;
> 4) the tendency toward a narrative mode of thought.(Stanovich, 2003, p. 6)

This points to serious problems not currently addressed in IQ research. Consumers of our products will often rely upon heuristics instead of careful analysis, and these same heuristics may lead to misjudgments with potentially large consequences.

Despite the wealth of research demonstrating the numerous fundamental errors humans can make when relying upon heuristics, such as the Jack, Anne, and George problem presented above, not all researchers accept them as demonstrative of flawed human decision processes. They believe the errors are instead with the researchers themselves, that construct hypothetical situations with nuanced language that will naturally confound something as complex as human rationality, that can not only solve rational problems, it has the ability to go far beyond just the information given, such as that needed for understanding metaphors and poetry. Gigerenzer protests that a research question such as the Jack, Anne, and George problem is "content-blind" and which

> …eliminates the characteristics of human intelligence from the definition of good judgment…. As a consequence, we have learned nothing about the nature of thinking or other cognitive processes on content-blind norms. Inappropriate norms are not simply a normative problem. They tend to suggest wrong questions, and the answers to these can generate more confusion than insight into the nature of human judgment. (Gigerenzer, 2005, p. 209)

However, Gigerenzer's critique of the research surrounding cognitive problems associated with the use of heuristics is easily countered by evidence where lapses of human rationality have lead to devastating results. In such examples, we are presenting the case for improving human rationality much like many practitioners of IQ present the consequences of poor data quality as a reason to initiate quality improvement programs. Such examples of bad decisions despite the existence of good data are the reciprocal of poor decisions resulting from poor data.

A good example of the consequences associated with heuristic processing instead of careful reasoning of the information presented would be the *Vincennes* incident. Curiously enough, this same incident is used to demonstrate the effects of poor IQ. From a cognitive psychology perspective, the situation is easily explained--even when presented with the same data, different consumers of data may arrive at different conclusions, depending upon the context of their perspective, that is, the task environment from which the decision is being made. From an IQ perspective, the same incident is much more difficult to explain, unless of course key elements of the incident are withheld.

On July 3, 1998, two U.S. Navy skippers, Captain William Rogers of the *U.S.S. Vincennes* and Commander David Carlson of the *U.S.S. Sides*, were evaluating a radar track from a plane that had taken off from Bandar Abbas, an Iranian airfield that served both commercial and military aircraft. Eighteen nautical miles separated the two ships, but both skippers had the same "view" of the theatre of operation thanks to the Aegis combat system that linked the two ships together. Despite this shared consistent view of the data, an important IQ feature, the two captains came to far different conclusions about the plane, assigned tracking number (TN) 4131 by the Aegis system.

The *Vincennes* had designated TN 4131 as "possible F-14" (Iranian fighter jet), but on the *Sides*, Carlson dismissed it as a possible threat:

> I evaluated track 4131 verbally as not a threat. My TAO gave me a quizzical look, and I explained. "He's climbing, He's slow. I don't see any radar emissions. He's in the middle of our missile envelope, and there is no precedent for any kind of an attack by an F-14 against surface ships. So, non-threat" (Evans)

We can see that Carlson approached the decision in a careful, rational manner and was able to correctly surmise the situation, despite the troublesome point that it could be a possible hostile fighter. At the same time, the *Vincennes* was in a heated surface engagement with a group of small, Iranian gunboats that had been harassing a merchant ship in the area. Rogers informed Fleet Command of his intent to engage TN 4131 at twenty nautical miles if it did not turn away, but at least one of his own crew expressed concerns about the target to Rogers:

> The CIC officer, Lieutenant William Montford… saw a mode III (civilian aircraft) at an altitude of 8,000 or 9,000 feet and rising slowly. He stepped forward and said, "Possible commair [commercial airliner]." I extended an arm over my head and acknowledged him. (Rogers & Rogers, 1992, p. 14)

When TN 4131 was within ten nautical miles of the *Vincennes*, still engaged against the Iranian gunboats and in a full-rudder turn at 30 knots which spilled books and equipment from the racks, Captain Rogers gave the command to launch two SM-2 antiaircraft missiles, despite the clear warning from Lt. Montford. There were no survivors among the 290 aboard TN 4131, which was in fact Iran Air Flight 655 on a routine commercial flight to Dubai.

While it might be argued that the decision to launch an attack against a possible hostile target as a System 1 process, "threat" or "non-threat," the careful analysis of the situation by Cmdr. Carlson, repeated by Lt. Montford on the *Vincennes*, is demonstrative of the fact that a multitude of factors weighed in on their appraisal of the situation. Capt. Rogers took a much simpler route to arrive at a much different judgment.

If we are to measure the data inputs used in our systems as to their fitness for use, we must also attempt to determine the fitness for use of the systems itself in the task environment for which it was designed. How well does it support the decision processes with which it was created to serve? It is not sufficient to document that the data inputs are timely, complete, accurate, etc. In the case of the *Vincennes* incident, both the data inputs and the system itself were of sufficient quality for a correct decision, as evidenced by the correct interpretations of Cmdr. Carlson and Lt. Montford, This was obviously not the case for Capt. Rogers. The system itself was not of sufficient quality to prevent Capt. Rogers from falling into the trap of a fundamental computational bias.

In a famous article on IQ by Fisher and Kingma, Rogers' error is attributed to the reuse of tracking numbers which led the crew of the *Vincennes* to confuse the radar track of an American A-10 with a Iran Air flight 655 and its downing--if "the duplication of identifiers had been recognized, the involved parties could have avoided the disaster." (Fisher & Kingma, 2001, p. 113) This is a somewhat more satisfying answer to why 290 civilians were accidentally killed. It suggests that a simple fix could be put in place to prevent this from ever happening again. However, the formal investigation into the incident concluded that there were no such data irregularities as identified by Fisher and Kingma:

> The AEGIS Combat System's performance was excellent—it functioned as designed. Had the CO USS *Vincennes* used the information generated by his CAD system as the sole source of his tactical information, the CO might not have engaged TN 4131. (Department of Defense, 1988, p. 43)

The *Vincennes* incident and Captain Roger's actions are best explained by what Stanovich labels the fundamental computation bias, in this particular case, "the tendency to contextualize a problem." (Stanovich, 2003, p. 309) In explaining the moment the decision to fire was made, Rogers recalls that "I was now convinced, beyond doubt, that the aircraft was supporting the surface engagement in progress, and that my ship and crew were in imminent danger. (Rogers & Rogers, 1992, p. 16) This passage makes it clear that Rogers linked TN 4131 with the surface vessels with which his ship was engaged in battle. This link existed only in his mind and was not supported by data, which is clearly evidenced by Carlson and Montford's correct interpretation of the same data, despite the fact that the two men were eighteen nautical miles apart and that Montford was himself in the same heightened combat situation as Rogers. These computational biases are a product of our evolutionary past often at odds with the modern world, which can at times have devastating results:

> the modern world presents situations in which the type of contextualization rendered by the fundamental computational biases proves extremely problematic. Such situations are numerically minority situations, but they tend to be ones where a misjudgment tends to have disproportionately large consequences for a person's future utility maximization… (Stanovich, 2003, p. 294)

Despite the critique of Gigerenzer and others to the contrary, there are numerous examples, such as the *Vincennes* incident, where the reliance upon heuristics leads to terrible results. We can accept the fact that the human mind is capable of interpreting nuanced circumstances that are not captured by the framing of research questions designed to evaluate the limits of human rationality. However, there are still many examples where the reliance on heuristics, what we normally recognize as "gut instinct," are not only incorrect, they can lead to consequences that have considerable impact. Given this fact, if we are to evaluate the fitness of use for our IP system, we must be aware of how our system may be potentially misused by users, in the sense that our representations, although they may be complete, timely, and accurate, etc., can ultimately lead to bad decisions simply because their representation led to a cognitive overload for the consumers, and they used them in fundamentally irrational ways.

Of course, one approach to lessening the cognitive burden on consumers would be to simply limit the number of data inputs available for them. Gladwell advocates for this approach in his book *Blink*, recounting the success Cook County Hospital had when they introduced a protocol for the evaluation of patients suffering from chest pains which limited emergency room physicians to just check four key indicators, ECG, blood pressure, fluid in the lungs, and unstable angina:

> …extra information is more than useless. It's harmful. It confuses the issues. What screws up doctors when they are trying to predict heart attacks is that they take *too much* information into account. (Gladwell, 2007, p. 137)

It is easy enough to understand how limiting access to information for decision makers can lead to better decision outcomes. By limiting the amount of information available, one decreases the complexity of the task environment, helping the decision maker avoid relying upon heuristics and the computation biases associated with such an approach. However, we can also imagine a large number of instances where all available data is necessary for effective decisions, and organizations that spend a great deal of time and money maintaining their data assets would most likely want them deployed in an effective manner.

When all available data is needed for effective decision making, we need to find a different approach to lessening the complexity of the task environment, the cognitive load placed on decision makers, or both. While we can certainly not avoid misuse of our products, we can evaluate how consumers transform data into information with our products. By focusing on the system's use, we can apply our quality improvement initiatives toward making this transformation suitable for the task environment and improving our consumers' ability to process all available data.

To return to Simon's concept of bounded rationality, if we are to construct and manage IP systems for deployment in System 2 processes, we must concern ourselves with the task environment confronted by our consumers. The limitations of human computation of are of primary concern:

> The scarce resource is computational capacity--the mind. The ability of man to solve complex problems, and the magnitude of the resources that have to be allocated to solving them, depend on the efficiency with which this resource, mind, is deployed. (Simon, 1978, p. 13)

This is an important concept as we attempt to discern where we should apply quality improvement initiatives to the development and management of our products for System 2 decision processes. “If attention is the scarce resource of a decision maker, then helping individuals manage attention is critical for improving decisions.” (Payne & Bettman, 2007, p. 112) That is, if we wish to make IP systems as suitable as possible for complex task environments, then it is contingent upon developers to limit the cognitive load required by our consumers to access our products.

Amarel acknowledges the power an appropriate representation of the problem can have in finding a solution: “…by furnishing a man with convenient graphical displays of appropriate models, he will be stimulated to provide the creative contribution expected from him in his problem-solving…” (Amarel, 1966, p. 113) Simon echoes this sentiment by declaring that “solving a problem simply means representing it so as to make the solution transparent.” (Simon, 1996, p. 132) Both Amarel and Simon are referring to visualization as the representation needed to facilitate problem solving.

As Tufte points out, “Graphics reveal data. Indeed graphics can be more precise and revealing than conventional statistical computations.” (Tufte, 2006, p. 13) Perhaps more importantly, Tufte also provides us with an important metric in helping us determine the fitness for use when applying visualization to an IP system:

> Data graphics should draw the viewer's attention to the sense and substance of the data, not to something else. The data graphical form should present the quantitative contents. Occasionally artfulness of design makes a graphic worthy of the Museum of Modern Art, but essentially statistical graphics are instruments to help people reason about quantitative information. (Tufte, 2006, p. 91)

Ware suggests why data representation, specifically data visualization, should play a pivotal role in the creation of products for System 2 processes, especially given the realization of increasingly complex task environments:

> Visual displays provide the highest bandwidth channel from the computer to the human. We acquire more information through vision than through all the other senses combined… Improving cognitive systems often means tightening the loop between a person, computer-based tools, and other individuals. On the one hand, we have the human visual system… On the other hand are the computational power and vast information resources of the computer… Interactive visualizations are increasingly the interface between the two. Improving these interfaces can substantially improve the performance of the entire system. (Ware, 2004, p. 2)

In an increasingly complex world, where the amount of information the individual is expected to both process and act upon, data visualization provides the means by which we can reduce complexity and derive meaning.

If we return to our discussion concerning the relationship between data and information, we can see now that the role of data visualization is to aid the data consumer in the transformation process in which they turn data into information. Knowledge, whether it is explicit or tacit, provides the necessary context for this transformation process to occur. Data visualization is not the only way to facilitate this transformation, but as Ware points out, it is an efficient way to connect vast amounts information to the serial processing of pattern recognition represented by human cognition. Data visualization serves to present the consumer of products for System 2 decision processes information in the same serial processing mode such decision processes require.

The Nature of Knowledge as a Social Function

On August 31, 1854, a cholera outbreak began along Broad Street, in the Soho section of London. Prior to this time, the prevailing theory was that all diseases stemmed from "miasma," or bad air. Edwin Chadwick, who had the distinction of working as both commissioner of the sewers and of the General Board of health, testified before a parliamentary committee in 1846 that:

> All smell is, if it be intense, immediate acute disease; and eventually we may say that, by depressing the system and rendering it susceptible to the action of other causes, all smell is disease. (Johnson, 2006, p. 114)

Microscopes of the time generally did not have sufficient resolution to detect the bacterium responsible for cholera, *Vibrio cholerae*. An Italian anatomist, Filippo Pacini, was able to isolate the bacterium in 1854, but the prevailing belief in the miasma theory of the disease prevented Pacini's work from being recognized until many years after his death. ("Filippo Pacini", 2009)

A London physician, John Snow, had written a paper prior to the Broad Street outbreak, suggesting that the source of cholera was not foul air but instead the drinking water shared by the effected population. However, the theory of miasma persisted. When the outbreak along Broad Street occurred, Snow saw this as another opportunity to prove his theory. He cataloged the deaths of the outbreak into a map, showing their proximity to the pump on Broad Street.

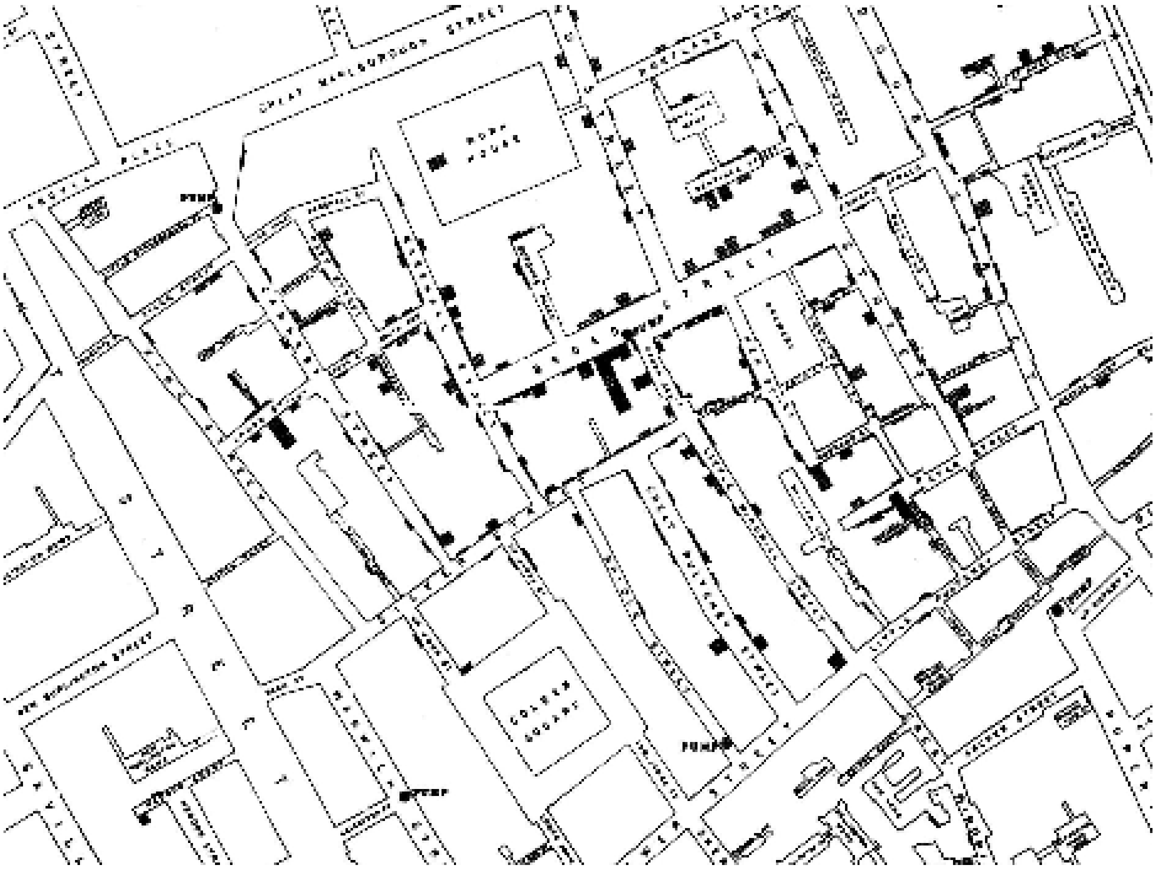


Figure 1: Excerpt from the original map made by John Snow in 1854. The Broad Street pump is visible in the center of the image, and cholera deaths are represented by black bars. (“The John Snow Archive and Research Companion“, n.d., figure 1)

Snow’s map was not at first accepted as the definitive proof of the link between cholera and drinking water, but other accounts of the outbreak used Snow’s map as an illustration. In time, this broad exposure to a brilliant visualization of data, a simple explanation that anyone could quickly analyze, helped to give the waterborne theory of cholera wide acceptance. Johnson’s *The Ghost Map* is both a retelling of this story and an investigation of how wrong ideas, such as the theory of miasma, are so persistent. Johnson credits Snow’s determination in proving his theory and the elegance of his methodology:

> And so the ghosts of the Broad Street outbreak were reassembled for one final portrait, reincarnated as black bars lining the streets of their

> devastated neighborhood. In dying, they had collectively made a pattern that itself point to a fundamental truth, though it took a trained hand to make that pattern visible. (Johnson, 2006, p. 197)

As with our previous examples of the fundamental computational bias, knowledge as a social function can exhibit a similar myopia. Kuhn posits that current scientific thinking represents a "paradigm" through which all research is filtered. The paradigm only changes when a practitioner discovers a case not fully explained by the paradigm, and the new paradigm is incommensurable with the previous:

> Normal science, the activity in which most scientist inevitably spend almost all their time, is predicated on the assumption that the scientific community knows what the world is like. Much of the success of the enterprise derives from the community's willingness to defend that assumption. Sometimes a normal problem, one that ought to be solvable by known rules and procedures, resists the reiterated onslaught of the ablest members of the group within whose competence it falls…. In these and other ways besides, normal science repeatedly goes astray. And when it does—when, that is, the profession can no longer evade anomalies that subvert the existing tradition of scientific practice—then begin the extraordinary investigations that lead the profession at last to a new set of commitments, a new basis for the practice of science.(Kuhn, 1996, p. 5-6)

As with the case of Snow's Ghost Map, simply showing even a large audience compelling evidence does not warrant a change in perception. It may often be the case that repeated demonstrations of the evidence, from different participants, will be needed before a new understanding is accepted and adopted. If our IP system is to be utilized for organizational knowledge, we must realize such knowledge is resistant to change, and new ideas will most likely require extensive exposure before they are accepted.

This is characteristic of the social nature of knowledge. It is resistant to change until that time when it can be demonstrated that existing knowledge does not fit our perceptions of the world. In this context, our inability to distinguish between data and information can have severe consequences, since it may lead us to model our world in data as we currently understand it instead of in ways that might illuminate the inadequacies of our current understandings:

> Knowledge is neither data nor information, though it is related to both, and the differences between these terms are often a matter of degree. We start with those more familiar terms both because they are more familiar and because we can understand knowledge best with reference to them. Confusion about what data, information, and knowledge are—how they differ, what those words mean—has resulted in enormous expenditures on technology initiatives that rarely deliver what the firms spending the money needed or thought they were getting. Often firms don't understand what they need until they invest heavily in a system that fails to provide it. (Davenport & Prusak, 1998, p. 1)

The problems Davenport and Prusak refer to are systematic of the approach where user needs and system specifications for the entire system or acquired first, and business value is delivered only at the end of development; the waterfall development model instead of iterative and incremental development. What is missing from such approaches is the simple awareness that user needs develop in step with organizational knowledge. Such an iterative approach to system development is an outgrowth of the "plan-do-study-act" cycles of quality improvement pioneered by Shewhart and Deming.(Larman & Basili, 2003, p. 47)

This calls into question the common notion that data is rolled up into information which itself is rolled up into knowledge. Tuomi contends that the reverse is actually true. It is current knowledge, our existing paradigm, that determines the information we consider important, and in turn, the data we choose to collect to support this information which supports our knowledge. We build information systems in support of this approach, and the data that we choose to store in them are the result of breaking down our knowledge into storable "bits."

> The meaning structure that underlies knowledge for an individual is articulated through cognitive effort to become focal and structured…. When such articulated knowledge is stored in computer memory for automatic manipulation, the meaning of information must be represented. In effect, information has to be split into “atoms” that have no meaning that would need to be taken into account in automatic processing. At this point we have created data.
> (Tuomi, 1999, p. 115)

Toumi’s “reverse knowledge hierarchy is reminiscent of Vygotsky’s rejection of the behaviorist model of speech development from egocentric thought and speech to socialized speech and logical thinking. For Vygotsky, the whole purpose of speech is social interaction, so it develops instead from something outside the child which is eventually internalized into inner speech the individual uses for thinking:

> We see how different is the picture of the development of the child’s speech and thought depending on what is considered to be the starting point of such development. In our conception, the true direction of the development of thinking is not from the individual to the social, but from the social to the individual.(Vygotsky, 1986/1986, p. 36)

This understanding of knowledge as a social function is critical as we design products for System 2 decision processes, especially of course for those that are also needed to expand organizational knowledge. We can evaluate decisions from a normative perspective, what decisions we believe individuals should have made, or from a descriptive perspective, how individuals actually make decisions. As noted before, the outcomes for many types of decisions, such as perhaps picking a particular stock or mutual fund to add to one’s portfolio, may not be clear for some time to come. In fact, many shareholders will often sell stocks or funds at the first hint of a decline in value, even though the selling price is actually lower than that which they paid originally. Investors clamor to buy a stock that suddenly becomes hot and quickly dump the stock once they sense their investment is going south. The problem with this approach of course is that investors are buying high and selling low. From a purely descriptive perspective, we can see the results of bad decisions in the stock market,

> …Firsthand Technology Value earned an impressive 16% annualized total return from the beginning of 1998 through the end of 2001. But we calculate that its typical investor lost 31.6% annually over the period. Overall, shareholders lost $1.9 billion… Technology funds did the most damage, vaporizing $30.5 billion of shareholders' money. At the Janus funds, the average portfolio returned 5% annually from 1998 through 2001, but the typical fund investor at Janus lost an annual average of 11.1%. That translates to $7.3 billion in wealth wiped out. (Zweig, 2002)

From a normative perspective, we could point to the benefits of a diversified portfolio and long-term thinking about investments.

Individuals are not the only ones prone to poor investment decisions. The Pew Center on the States estimates that at the end of fiscal year 2008, there was a $1 trillion gap between the liability of state pension funds and their actual value, and that only four states had fully funded systems. (The Pew Center for the States 2010, p. 1) One of the key elements missing from Huang, Lee, & Wang's requirements for an IP system created to manage organizational knowledge is a recognition that an important area for consideration would be for the individuals in an organization to learn from mistakes in order to be better prepared for the future. Somehow capturing this knowledge, both mistakes and successes, and promulgating what has been learned is a critical need as the system becomes more complex and changes become more rapid. For example, Gawanda points out that discoveries and advancements in medicine are so rapid that the systems used to help practitioners manage patient care is unable to capture it:

> The software used in most American electronic records has not managed to include all the diseases and conditions that have been discovered and distinguished from one another in recent years…. The complexity is

> increasing so fast even the computers cannot keep up. (Gawanda, 2009, p. 22)

To help medical professionals deal with the increasing complexity of medical knowledge and practice, Gawanda advocates for the use of a checklist, to ensure that nothing is missed when a diagnosis is being contemplated and a treatment is being considered. We can view this as an important alternative to Gladwell's assumption that less information is always better. Gawanda's approach is reminiscent of the TADMUS program, which the Navy initiated in response to the *Vincennes* incident. Both approaches focus heavily on the user's perspective. We can view both approaches as helping to enforce normative rules in decision making to facilitate well-defined descriptive processes and thus produce both better and more consistent outcomes. We must somehow constrain the dimensions upon which decisions are made because without them, individuals within the organizations will naturally have different perspectives on what particular dimensions are most important in any given decision process. Without such constraints, it is easy to imagine a situation where one member of the organization would imagine the slow, climbing radar contact on his system as "non-threat" while the other would imagine the contact identified as "possible hostile" to be an immanent threat to his ship and crew. Restricting the amount of data we provide to decision makers is much less satisfying than simply ensuring that they are well aware of the normative rules associated with the decision process.

While it is difficult to imagine a system that could prevent all individuals within an organization from making a poor decision, it is the interaction of individuals within the organization that defines the normative rules adopted for decision making. These rules serve an important role in that they make explicit knowledge within the organization that may be tacit and internalized within individuals. The expression of normative rules within the organization, in the form of tradition, implied values, and inherit practices also helps guide individuals through the process of knowledge creation. As Polyani says, for the individual "we can know more than we can tell…. we are attending from these internal processes to the qualities of things outside." (Polanyi, 2009, p. 4-13).

This ability to articulate in an explicit manner the organization's norms, values, and practices is a necessary limiting factor when the sheer amount of data available for decisions is often overwhelming, yet we may not know what we do not know—it will often be the case that the problem is being formulated as it is being solved.(Spence, 2007, p. 17) Capturing the organization's explicit knowledge not only facilitates problem solving, it also helps define the domain in which these problems will be defined. Without such a framework, it will often be the case that the problem is being formulated as it is being solved. Schon's concept of "problem setting" is just such an example where explicit knowledge provides normative rules by which to not only guide decisions but to also determine what may be wrong with our decisions:

> But with this emphasis on problem solving, we ignore problem setting, the process by which we define the decisions to be made, the ends to be achieved, the means which may be chose. In real-world practice, problems do not present themselves to practitioners as givens. They must be constructed from the materials of the problematic situations which are puzzling, troubling, and uncertain…. When we set the problem, we select what we will treat as the "things" of the situation, we set the boundaries of our attention to it, and we impose upon it a coherence which allows us to say what is wrong and in what directions the situation needs to be changed. (Schon, 1995, p. 40)

Huang, Lee, & Wang also recognize the importance tacit knowledge plays within an organization and the need to capture this into explicit knowledge that can be reused by the organization, a process they refer to as “’knowledge hunting,’ the process of collecting knowledge, harvesting the process of filtering, and hardening the process of structuring tacit, useful knowledge into explicit, reusable knowledge.” (Huang et al., 1999, p. 114) This represents an important dimension as we seek to develop IP systems in support of System 2 process that also serves to help an organization’s knowledge management. We must facilitate the process of capturing discovered knowledge and turning this into normative rules that individuals within the organization can follow.

Chapter 3

Data Driven Decision Making in Education

The Task Environment of Education Data

Decisions related to education often require System 2 processes. For example, assessment data is “highly dimensional,” with multiple attributes for subtests and numerous categories of demographics, all of which need to be studied when making determinations as important as a child’s education. No Child Left Behind (NCLB) has made schools and districts accountable for their students’ proficiency in particular subjects and grades, with the goal of having all students proficient in mathematics and literacy by 2014. (“NCLB“, n.d.) NCLB has helped push the importance of data analysis for educators, and the federal government has also provided additional funding for states to create data systems specific to helping educators use data for program improvement. However, there is little guidance from the federal government on best practices related to analyzing student data to improve student outcomes.

The National Center for Education Statistics’ (NCES) Institute of Education Sciences (IES), in its Statewide Longitudinal Data Systems (SLDS) Grant Program, will have awarded $500,000,000 to states, territories, and the District of Columbia once its third round of funding is finalized in 2010. These grants are intended to help “states, districts, schools, and teachers make data-driven decisions to improve student learning.”(“IES SLDS Grant Program“, n.d.) In its “Forum Guide to Decision Support Systems,” the NCES outlines a general approach on how an SLDS should be constructed, but only a single page is devoted to user training, and this discussion is entirely devoted to planning for training and not specific in any way to best practices around data use.

(National Center for Educational Statistics [NCES], 2006) The Data Quality Campaign (DQC) has created a brief on data use, recognizing the need to transition from a focus on data for compliance to using data for continuous improvement. (Data Quality Campaign [DQC], 2009) Once again, this brief contains only general statements and no specific details on best practices to ensure consumers use data effectively.

The federal government, through the annual student assessments and accountability mandated by NCLB, has put data on the agenda of every educator in the country. Besides the accountability under NCLB, funding for federal grants and assistance begins with the requirement for schools, districts, and even state agencies to begin their requests with an analysis of their current data to determine specific needs. The IES SLDS grant program, a $500,000,000 commitment from the federal government, demonstrates that this insistence on data driven decision making (DDDM) is also well funded. However, schools, districts, and states are currently left to their own devices to determine what processes they might deploy for effective analysis of data.

While there are many criticism of NCLB, the lack of guidance concerning DDDM is particularly problematic, especially since so many of the decisions educators must make are clearly System 2 processes. Educators' interventions in programs and instruction to improve the number of students reaching proficiency may not have a measurable impact for some time, and the full impact of their interventions may go unnoticed by educators as they focus on the results to a particular problem without realizing their "solution" has actually caused problem in other areas. An example of this would be a school's decision to give additional resources to "bubble-kids," students that were very close to proficiency under NCLB. This is a common practice in schools across the nation. The idea is that by focusing limited resources on students closest to proficiency, the school has the best chance to meet its accountability goals on the next exam. What this strategy fails to address is that a whole host of students will grow further behind their peers because they are not receiving a similar focus. It also fails to recognize that just because a student was determined to be proficient under NCLB one year does not mean that they will necessarily stay proficient without proper interventions the next. Without proper guidance, educators will often resort to heuristics in response to the overwhelming task environment related to System 2 processes. As Simon noted, when confronted with unfamiliar domains, "people satisfice—look for good-enough solutions--instead of hopelessly searching for the best solution." (Simon, p. 17)

Accountability under NCLB has made DDDM something of a mantra in education, but there are many unanswered questions about how educators should interpret and analyze data and about the effects DDDM is having on educational outcomes (Marsh, Pane, & Hamilton, 2006, p. 1). There are some general models for implementing DDDM, most modeled after Deeming's Plan, Do, Study, Act cycle. For example, Harvard's "Data Wise" improvement process lists eight steps, under the general categories of Prepare, Inquire, and Act (Boudett, City, & Mumane, in press). Victoria Bernhardt's model, Education for the Future, is a bit closer to Deeming's original model with seven steps under the categories of Plan, Implement, Evaluate, and Improve, but these steps are organized around another central category of Vision (Bernhardt & Geise, March 15, 2009, p. 9). While both of these models represent the current thinking on how best to use data in education, neither offer any specifics on how to implement these programs or any metrics to judge the efficacy of the approach in improving student outcomes. Both programs represent instead general guidelines to consider when analyzing education data. It is also important to note that none of these models give any considerations to the mistakes consumers may make, especially when confronted by System 2 decision processes.

More comprehensive models of DDDM have been proposed. Mandinach, Honey, and Light propose a framework whereby individuals at different levels of the hierarchy have questions or problems which require data to be collected and analyzed in order to make informed decisions (Mandinach, Honey, & Light, April 9, 2006). Ikemoto and Marsh expand on this framework to differentiate both "simple and complex data" and "simple and complex analysis and decision making" (Marsh et al., 2006, p. 113). Ikemoto and Marsh are somewhat dismissive of the Mandinach, Honey, and Light framework, pointing out that "it fails to capture the nuances and variation that occur when educators go about making decisions in real-world settings," and that "DDDM in practice is not necessarily as linear or continuous as the diagram depicts" (Marsh et al., 2006, p. 110). What Ikemoto and Marsh fail to recognize is that even "simple analysis and decision making" is fraught with potential problems, specifically, the fundamental computational bias. Even "simple" analysis and decision making can lead the data consumer down erroneous paths if the common errors of data analysis are not first recognized and somehow accommodated. While the consequences of poor data analysis in an educational setting may not be on par with the consequences associated with labeling a commercial airliner as "hostile" in a combat situation, it cannot be stressed enough that the consequences for a child labeled "not proficient" are indeed quite significant to that child, his/her parents, his/her school, and a whole host of stakeholders in the K12 educational chain. In aggregate, the mistakes we make educating our children eventually impact both the economy and society of our nation.

One problem often missing from the literature surrounding DDDM in education is that the task environment is very challenging. Both the data and decisions are very complex, even when they may appear simple. Education data is often highly dimensional, the problem setting is seldom well defined, and often the full results of decisions will remain hidden. If we examine the other blade of Simon's scissors, computational ability, we find that the deciders, educators, are woefully unprepared for these complex decisions and task environments. Collecting data and aggregating it in meaningful ways requires technical skills that are somewhat arcane and certainly not a focus of teacher and administrative preparation. Even the most rigorous of training in the technical and analytical skills needed for data analysis of multi-dimensional data would not necessarily provide one with the resources needed to make sound decisions from educational data. Even at the classroom level, looking at results of a single assessment means evaluating the responses of twenty to thirty independent actors, students, for each assessment, not to mention individual responses for each question.

When decision tasks require a tremendous amount of computational ability, the default response is to fall back on heuristics, which Stanovich points out "do not permit fine-grained accuracy, but they are fast acting, do not interfere with other ongoing cognition, require little concentration, and are not experienced as aversive" (Stanovich, 2009, p. 63). While consumers of education data may attempt a sincere and thorough analysis before reaching a decision, the complexities of the data and the educators' own limited skills and training in data analysis means that they will rely much more heavily on heuristics than rationality. To understand the full implication of this reliance upon heuristics, gut-instinct, instead of well-reason rationality to children, consider that a teacher's appraisal of a student can often lead to what Rosenthal and Jacobson label the "Pygmalion effect"—"one person's expectation for another person's behavior can quite unwittingly become a more accurate prediction simply for its having been made."(Rosenthal & Jacobson, 2003, p. vii) A teacher's gut instinct about a student, the teacher's reliance on heuristics instead of careful reasoning, can easily label a student as "low achieving," and the Pygmalion effect will tend to make that a self-fulfilling prophecy for the student thus labeled.

Any approach to improving the use of data for DDDM in education must deal with both boundaries of the problem, the task environment and computation ability, while somehow keeping in check the tendencies to rely on computational processes that sacrifice accuracy and granularity in the interest of ease and speed. Unfortunately, IQ as a discipline has not reached a level of maturity where we can "close the gap" between the user's decisions and the IP system provided. At this point in time, the closest tool we have to assessing users' impressions about the usability of the data would be the Data Quality Assessment tool developed by Pipino, Lee, and Wang (Pipino, Lee, & Wang, 2002). It is important to note that this assessment tool does not evaluate decisions made by consumers of the information provided, but it does at least provided some objective measure of users' perceptions of the data presented and what they believe is their ability to act upon these data. Previous use of the Data Quality Assessment tool has shown results from the tool to be consistent among a diverse group of educators evaluating dissimilar data available from the same general system (Gibson & Decker, 2006).

One of the shortcomings of the Data Quality Assessment tool is that it does not query users concerning enhancements they believe might be important in improving the system being assessed, that is, it does not provide us with the proper metrics for iterative development. It is instead a measurement of perceptions concerning quality dimensions of the data used in the system. It also fails to accommodate for what mistakes a consumer of an IP system might make in his/her application of the data presented, regardless of the data dimensions' quality. For System 1 processes, where the outcomes are readily apparent, this approach is adequate. A poor outcome from a decision can often be related directly to a particular dimension, such as the timeliness or accuracy of a direct-deposit message. For System 2 decision processes this approach is much less satisfying. A user may be unaware of potential data that could be made part of the IP system and be a significant aid in their decision making. An example of this would be the addition of individual student growth data that can help educators spot potential problems for students that have seemingly high scores but which may not have received the necessary growth for the next level of instruction.

To actually improve the systems associated with System 2 processes, we do not have quantifiable instruments such as the Data Quality Assessment tool. At this point in time, we must instead emulate the TADMUS approach and concentrate on user factors. We must actually witness the decision processes our consumers use with our systems, monitor the decisions made, and appraise whether or not these decisions match normative rules. In the case of IP systems for System 2 decision support that is also used for knowledge management, it also must be determined if there is a process for collecting explicit knowledge and to facilitate this processes. This does not diminish in any way the importance of the data dimensions assessed in the Data Quality Assessment tool and to ensure that the inputs used for the creation of our system are fit for use. However, to improve IP systems used for System 2 decision processes requires us to continually assess how our consumers transform the data presented into information and the normative rules that provide the context that guides this transformation. It must also aid in the acquisition of knowledge that will help form and adapt the normative rules for the problem domain.

Measuring Student Growth

Given the already highly-dimension nature of education data, it may come as something of a surprise that yet another dimension is necessary to provide educators with actionable data. However, an end of year assessment is not specific to what is learned by a student in that given year; it is also dependent upon prior learning that child has received. If a student's mathematics score is low in the eighth grade, there is a good chance his or her score was also low in previous years. Measuring student growth, that is the increase in learning for a given year as measured by the difference between the current year's score and the prior year's score, is more specific to a single year's learning. Adding a dimension for individual student growth that can be aggregated at the district, school, and teacher level represents another way educators can evaluate the strength of their programs and instruction.

With the advent of NCLB, states have been collecting annual assessment data for the majority of K12 students and using these data for school accountability. One of the frequent complaints about NCLB is that it treats all schools the same. A school with a relatively affluent student population is held to the same standards as schools that may have special challenges, such as high rates of poverty or large numbers of English language learners. While it is a laudable goal that all students become proficient regardless of their location, schools with predominantly large populations of poor and English language learners are much more likely to be subject to sanctions than schools with more affluent student populations. In response to these criticisms, the U.S. Department of Education encouraged states to submit proposals for incorporating growth models for accountability reporting under NCLB.

Since the accountability under NCLB is primarily directed at schools, the growth models proposed by states use various means to give schools credit for students that did not reach proficiency but which had demonstrated acceptable progress toward reaching proficiency by 2014. For example, the Arkansas growth model takes an individual student's current score and the score needed to reach proficiency at the eighth grade. A trajectory is created for each student that determines the requisite growth each year for the student to reach proficiency in eighth grade. As long as the student meets this requisite growth each year, the school can count them as part of their total number of students meeting adequate yearly progress. ("Arkansas growth model proposal", 2006, p. 9)

Arkansas' growth model is termed a "Growth to Proficiency" model, which is the most popular model among states that have implemented growth as part of the accountability under NCLB. ("Guide to United States Department of Education growth model pilot program", 2009, p. 14) Colorado also uses a Growth to Proficiency model, but their methodology also produces a "Student Growth Percentile" (SGP) for each student that is then aggregated at the school level. This provides Colorado with another dimension for determining school quality other than just the number of students reaching proficiency and those on track for reaching proficiency. ("The Colorado growth model: Higher expectations for all students", 2008, p. 9) However, Colorado is not able to aggregate data at the teacher level, since they do not capture school schedules. Arkansas does capture scheduling information for K12 students. Since the SGP provides growth information for each individual student which can be aggregated at the teacher, school, and district level, this information could be valuable to a variety of stakeholders in Arkansas and provide them more data that would be useful for DDDM.

It may seem obvious that if states are testing students every year, then a simple measure of growth would be the difference between a student's assessment score in the current year relative to the same student's score in the previous year. For example, if Noah scored a 702 on the state's fifth grade math test in 2008 and a 775 on the sixth grade math test in 2009, the difference would be 73 points. In contrast, if Ben scored a 425 on the fifth grade math test in 2008 and a 527 on the sixth grade math test in 2009, the difference between those two tests would be 102. Can we say that Ben "grew" more since the difference between the two tests was greater?

The problem with this simplistic approach to growth, which is used by many educators to estimate growth, is that Noah's scores on both tests are very high, while Ben's scores for both tests are relatively low. There is less potential for Noah to score higher, since he is already near the top of student scores, while Ben would need to score much higher before he even reaches proficiency, much less begins to match Noah's scores. It is perhaps even more simplistic to rely solely on NCLB proficiency, that is, simply count the number of students that moved up in proficiency, say from Basic to Proficient or from Proficient to Advanced. Noah's score in the sixth grade was sufficient to say he is already proficient at the seventh grade level, even before he begins classes as a seventh grader. The nature of the cut scores for proficiency is also problematic. A student can miss being proficient by a single question. Is there really a significant difference between two students that are only one question apart to determine that one is proficient and the other not?

If we wish to use growth as a measure of school or teacher effectiveness, we are presented with a new set of problems. It is recognized that accountability under NCLB is somewhat flawed since it is only concerned with a student's performance at a given point in time; it does not account for where a student was prior to beginning instruction for a given year. This is what has given rise to the acceptance of growth models, to build in some flexibility for accountability. Another problem arises because students are not randomly assigned to either teachers or schools. A student is routinely placed in a particular school based on his or her neighborhood, and parents will often advocate for their child to be placed with a particular teacher, defying a school's attempt to distribute students randomly among the available staff. In a similar fashion, teachers may use their seniority to choose which classes they wish to teach or the schools within a district in which they would like to work. Since our subjects are not randomly assigned, our ability to classify a teacher or even a school as "ineffective" based on proficiency, growth, or any other metric is problematic. All we can really say is that given a particular set of inputs, student scores over a period of time, a school or a teacher seems to be ineffective.

However, schools are not at liberty to decide which students they accept, since they are legally obligated to educate all children that live within their jurisdiction, and additional laws require that all students receive some form of education, the majority of which will be served by their local public school. Schools and teachers in poorer areas will continue to struggle in comparison with schools and teachers in relatively affluent areas, even given the application of growth models. The students from poorer areas still have to achieve the same proficiency as those in the wealthy areas, it is just that a growth to proficiency model may provide them more time to get there. Given the fact that many of the poorer students begin their public education far behind their more affluent peers, this extra time may certainly not be sufficient, and indeed, the actual increase in scores such students have to make each year to reach proficiency will be much greater than students that begin their testing already proficient.

If we could instead calculate a linear value for growth, such as that presented by SGP, aggregates of those percentiles would give us another dimension by which to judge the effectiveness of schools and teachers. Students will not be assigned to either schools or teachers randomly, so a single year's worth of data would be insufficient criteria to establish effectiveness, but if we were to provide multiple years of growth at the district, school, and teacher level, we could begin to see trends that would be strong indicators of effectiveness. If a school or teacher with a particular student makeup exhibits high growth over time, then a school with a similar student population that shows historically low growth might be compelled to visit the school or teacher with higher growth and see what is different about their programs or instruction.

Having a measure of individual student growth that can be aggregated at different levels would provide educators and other stakeholders with valuable data concerning the decision that might be made about the education of children. Consider a parent trying to decide which school they should choose for their child. Would it be better to send him/her to a school with overall high scores or high growth? The former would be a good indicator of student demographics, especially whether the student populate tends to be poor or affluent, but the latter is more descriptive of the actual teaching and learning going on at that school. Such a measure could also be used by the school to determine which teachers are more effective and to target professional development for teachers that the data suggests are less effective.

There is a growing awareness of the importance of calculating student growth. For example, states that participate in Phase 2 Stabilization Funding have to indicate whether or not individual student growth data is provided that is timely and informs instruction. (U.S. Department of Education, 2009, p. 2) Applications for Race to the Top must commit to an evaluation system for teachers and administrators that take into account individual student growth. (U.S. Department of Education, 2009, p. 9) This focus on individual student growth from the U.S. Department of Education is significant on two fronts. The first is the fact that the majority of growth models currently authorized by the U.S. Department of Education under NCLB are not suitable for this type of application since they are growth to proficiency models instead of an individual calculation for each student. States will have to redesign their growth models for these particular applications. Given the fact that Phase 2 Stabilization Funding represents $1,400,000,000 and Race to the Top represents $4,000,000,000, there will be incentives for states to do this. The second point is that this is the first time the U.S. Department of Education has mandated teacher and administrator evaluations, preferring instead to simply hold schools and districts accountable. This is demonstrative of a growing awareness that evaluating teachers by student growth is an important dimension for DDDM.

Chapter 4

Creating a New Information Product for Education Data

Iterative and Incremental Development of hive

Improving IP systems associated with education data provides an excellent opportunity for research in improving systems for System 2 decision support. There is a wealth of student information available to educators, so much in fact that their ability to process it all is somewhat limited. As with the doctors at Cook County Hospital, a strong case could be made that educators have access to too much data, and that better outcomes might be achieved by limiting their access. Education data is often highly dimensional and just as often without additional information to put it in proper context. Educators often feel overwhelmed by data and there is little professional development specific to helping them become more proficient and comfortable with analyzing data. For example, there are hundreds of professional development modules available to Arkansas educators through the online portal IDEAS, including topics on how to create databases and use spreadsheets to organize data, but there are no modules specific to helping educators understand education data and making data driven decisions. (Arkansas Education Television Network [AETN], n.d.)

The Arkansas Department of Education (ADE) received a grant in 2006 for the creation of an SLDS from the IES. This was a very successful project, and Arkansas was acknowledged as being a leader in the use of data by educators. For example, Arkansas was one of the first states to implement “Ten Essential Elements” identified by the Data Quality Campaign necessary for an effective SLDS. (Data Quality Campaign [DQC], n.d.) One of the key things that differentiated Arkansas from other SLDS programs was that a major focus of the program was to deliver individual student data down to the teacher level, in a way which was both secure and compliant with all privacy laws.

ADE's SLDS program attracted the interest of researchers from across the nation. The Assessment and Accountability Comprehensive Center (AACC) approached ADE with a project specific to data use by educators, to "build capacity among administrators, teachers, and staff to increase the use of school-based data for improved student learning." (Heritage, 2009, p. 1) The objectives of the project were to develop professional development for educators on data analysis, help educators understand the questions that can and cannot be answered with current data, and identify improvements needed for ADE's SLDS.(Heritage, 2009, p. 1) The research was focused on two areas of the state, the southeast and southwest, and two districts from each area were selected to participate in the program.

Four separate sessions throughout the year were held with educators from these districts. The sessions progressed through data of increasingly granularity. The first session was devoted to district and school data, the second session was concerned with an analysis of strand data, the third session focused on individual student data, and the final session was specific to data collected and managed at the local level. In each session, elements of the fundamental computation bias were clearly evident. Educators would identify a single data point as indicating that there was a problem with a newly adopted textbook, a change in scheduling or classroom configuration, or a new program initiated by the district. Conversely, educators were also just as quick to contextualize a single data point as proving the efficacy of a program change they had made. Participants in the program were seen as continually engaging in a narrative mode of thought, as if all the data were trying to tell them something meaningful.

It was determined through these sessions that the data available through ADE's SLDS were not suitable for the types of analysis educators needed. While the SLDS did provide a significant breadth of data, the context for understanding these data was often missing. Participants had no way to compare district or school data to other districts or schools or even aggregate values for the state. Individual student data was available, but other than rudimentary aggregates of these data, participants had to manipulate these data themselves by first downloading the data and then importing it into a spreadsheet program. This was beyond the abilities of most participants. Participants did understand the importance of measuring student growth, so regression analysis was added as a topic. However, participants found the process of regression analysis far too difficult to be practical. ADE may have had what was considered to be an excellent SLDS, but in actual use, it did not represent a quality information product system with which users felt comfortable making educational decisions. It was evident that despite the praise Arkansas had received for its SLDS, its fitness for use was somewhat lacking.

The existing information systems used for these sessions came primarily from two sources, NORMES and Triand. NORMES is the vendor chosen by ADE for NCLB accountability. (http://normes.uark.edu/) As part of this mission, NORMES also provides educators with access to state, district, school, and individual student data, but the site is complicated for users and requires them to go to multiple sites for context:

District: BALD KNOB SCHOOL DISTRICT **Grade:** 7th

Select a grade:- 3 | 4 | 5 | 6 | 7 | 8 | 9| Algebra | Geometry

Combined Population

Year	Below Basic	Basic	Proficient	Advanced	Students Tested
2007	17.2%	14%	37.6%	31.2%	93
2008	17.7%	9.4%	37.5%	35.4%	96
2009	5.8%	14%	45.3%	34.9%	86

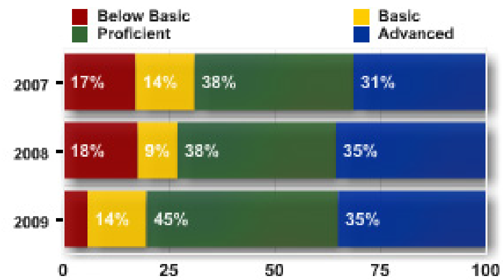


Figure 2: NORMES data visualization for grade seven mathematics at the Bald Knob school district. To see Arkansas averages for grade seven mathematics requires users to go to a different sites, as would comparing two different districts.

Another important provider of education data is Triand. (http://my.triand.com/) Triand is the vendor for ADE's electronic transcripts, but it also provides access to data at the state, district, school, and student level.

BALD KNOB SCHOOL DISTRICT - ACTAAP - 7th Grade - Mathematics

District	Administration	N	Mean Raw Score	Below Basic (BEL)	Basic (BAS)	Proficient (PRO)	Advanced (ADV)
BALD KNOB SCHO	Spring 2009	88	38	8.0	14.0	45.0	33.0
BALD KNOB SCHO	Spring 2008	102	40	19.0	11.0	37.0	33.0
BALD KNOB SCHO	March 2007	99	43	19.0	12.0	37.0	31.0
BALD KNOB SCHO	March 2006	97	34	27.0	16.0	41.0	15.0
BALD KNOB SCHO	March 2005	107	35	31.0	30.0	31.0	8.0

Figure 3: Triand data visualization for grade seven mathematics at the Bald Knob school district. To see Arkansas averages for grade seven mathematics requires users to go to a different sites. Users are able to only see their district's data in Triand.

While Triand does allow users to view aggregates of state data, it does not provide users access to other districts' aggregate data. Triand was developing visualizations for the data but dropped development on this to create a new user interface for their product. One other thing to note between NORMES and Triand is the number of students tested each year is not consistent between the two products. This is because NORMES excludes highly-mobile students, students not enrolled in the district as of October 1 the academic year of test, while Triand includes all students tested.

Despite the limitations of both these vendors' systems for educators, Arkansas has received great praise for its data use, which these two products play a key role. For example, the U.S. Chamber of Commerce, in its "Leaders and Laggards" report, gave Arkansas an "A" for its education data initiatives, one of only seven states to receive this grade. (U.S. Chamber of Commerce, 2009, p. 59) In a recent report from the Data Quality Campaign, Arkansas is listed as second in the nation for taking the most action the Data Quality Campaign feels necessary to ensure effective data use. (Data Quality Campaign, n.d.) That these systems did not provide data in a way that facilitated DDDM for users is demonstrative of the lack of research concerning practical applications for DDDM with education data.

To address the problems identified with ADE's existing IP systems, Simon's model of bounded rationality was applied, and problems were addressed by attempting to increase the computational ability of the actors while at the same time lessening the cognitive load they faced when analyzing education data. A framework and set of protocols for DDDM was developed based on previous work that had been successful in other areas. (von Houten, Mlyasaka, Agullard, & Zimmerman, n.d.) This framework is in constant revision based on user inputs, but the current version is available in Appendix A. The purpose of the framework is to increase the user's ability, by taking them through a stepwise process for analysis. As such, the framework serves the same purpose for which Gawanda advocates for the use of a checklist in hospitals—to ensure that all necessary steps are taken and avoid potentially costly shortcuts.

To support this framework, a custom visualization tool was needed that would help users answer each of the questions asked in the framework. The most critical feature for these visualizations would be different levels of context that would allow users to compare their district or school to other districts and schools, as well as the ability to compare different student populations, such as ethnicity and economic status, with each other. Making such visualizations interactive decreased the cognitive load for data analysis in comparison to existing systems. The feedback from the framework and visualizations used during this project helped define the needs of users. It also provided the base for the tools that were eventually used in the research for this dissertation to evaluate how improvements to IP systems in support of System 2 decision process should be approached.

As noted previously, DDDM in education is complicated by the highly dimensional nature of the data, the poor computational abilities of its decision makers, and the fact that the full effect of outcomes for decisions may be abstruse. Further complicating this task environment is that the problem setting is far from well defined, other than the general idea that educational data should lead to better student achievement. Based on the project with the AACC, it was determined that to improve ADE's IP system for DDDM, a data visualization tool should be created that provided users with the necessary background to compare a district, school, or individual student's data in a variety of contexts. This visualization tool should also serve to limit the cognitive load such data analysis represents, so consumers would be less likely to depend on heuristics and be more rational in their analysis. The data analysis framework should be modified to be specific to this visualization tool and data, to provide a step-wise process for a thorough analysis while avoiding the fundamental computation bias. Finally, the system should also help capture and promulgate knowledge discovered through its use, to help develop better normative rules for making education decisions in Arkansas. Again, the framework was developed to increase the computational abilities of the users, and the data visualization system, now called "hive," was created to lessen the burden associated with data analysis.

Such a focus on the needs of users and human factors is somewhat new to the IQ discipline, where the fitness of use for a system is assumed, given that the data inputs are of suitable quality and the processes for the system's creation are well documented and followed. Until we can develop tools for measuring the effectiveness of decisions made with IP systems from domains as complex as educational data, we must engage the users of these systems for System 2 decisions processes, to monitor their use and continually improve the system based on these dimensions of decreasing their cognitive load, providing the necessary context to aid in the transformation of data into information, and capturing and disseminating knowledge that can in turn be used to help shape the normative rules associated with DDDM.

A number of products were researched for the visualization tools that might be used to create the visualization tool for the new IP system. The general requirements were that it be open-source in order that it could be more easily modified to incorporate new visualizations and that it include the ability for social networking as the foundation for organizational knowledge creation and management. The prefuse information visualization toolkit met these two requirements, and the source code was readily available. (http://prefuse.org/) Prefuse was developed in Java, with which the researchers were already proficient. The developers of prefuse had already developed an application that incorporated social networking with data visualization, "recasting visualizations as not just analytic tools, but as social spaces." (Heer, Viegas, & Wattenberg, 2007, p. 1029)

While the prefuse source code included many different kinds of visualizations, the only existing visualization suitable for education data was a scatter plot. A data set of Arkansas education was prepared for testing with prefuse, and the first results were promising:

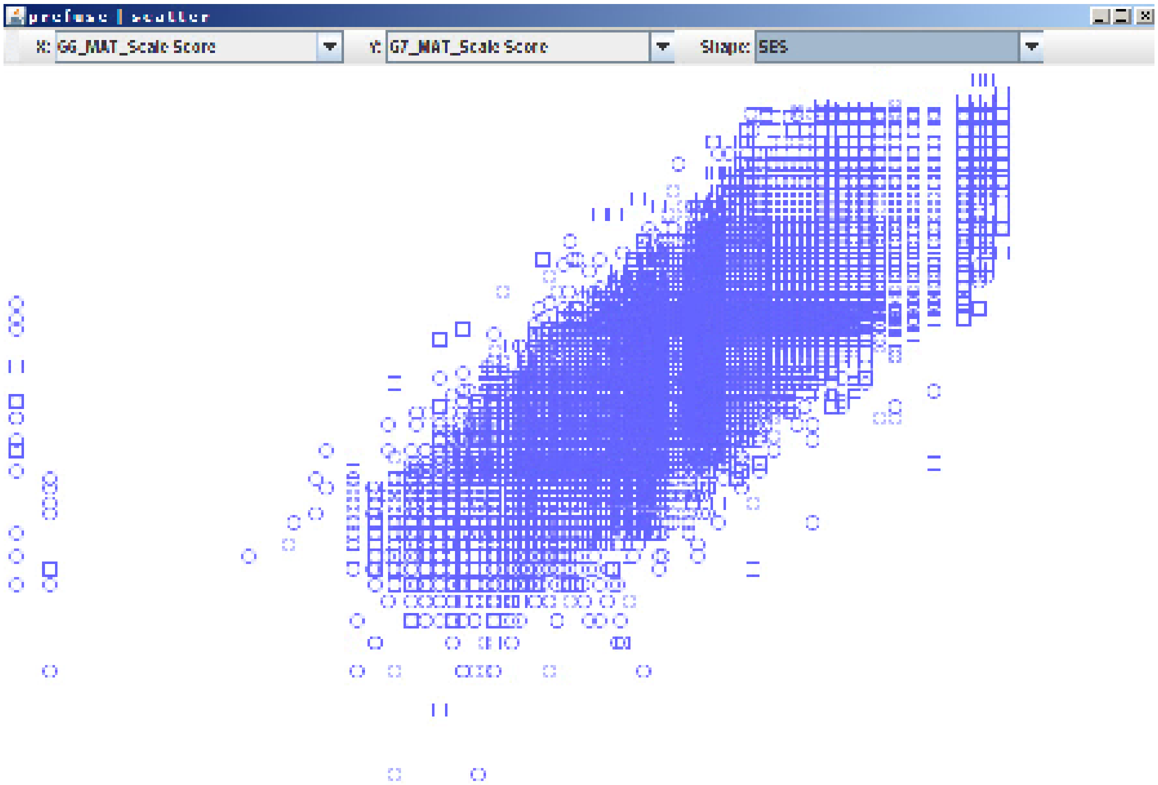


Figure 4: Prefuse scatter plot visualization of sixth grade 2008 mathematics scores with seventh grade 2009 mathematics scores. The different shapes of the plots (squares or circles) denote plots (individual students) that have the "shape" attribute, in this case, circles for those labeled as "economic disadvantaged," and squares for "not economic disadvantaged."

While this scatter plot offered basic functionality for use with education data, it was missing two key features that would extend its applicability. It needed the ability to display individual student names for when an educator identified a data point of interest, and it needed the ability for brushing and selecting on attributes.

Brushing is a powerful visualization technique in that it can be used to change the encoding of a particular element, allowing users to quickly discover groups of interest. For example, brushing can be used to quickly identify a particular population within a scatter plot of individual student scores by turning all the data points of a particular group, such as economic disadvantaged, to a different color. Adding to this is the concept of selection, where a group of interest can be chosen as the only data to be displayed. Quickly switching between brushing and selection allows a user to navigate quickly through various views of the data, confirming or disproving various concepts they may have of student performance. The existing prefuse toolkit had a similar methodology with the incorporation of "shape," but shape as an encoding was somewhat limited when large amounts of data were involved. A brushing and selection class were added to prefuse, which greatly increased its usefulness with large data sets. A mouse-hover class was added to the existing prefuse toolkit, which allowed for the discovery of individual student names within a data set.

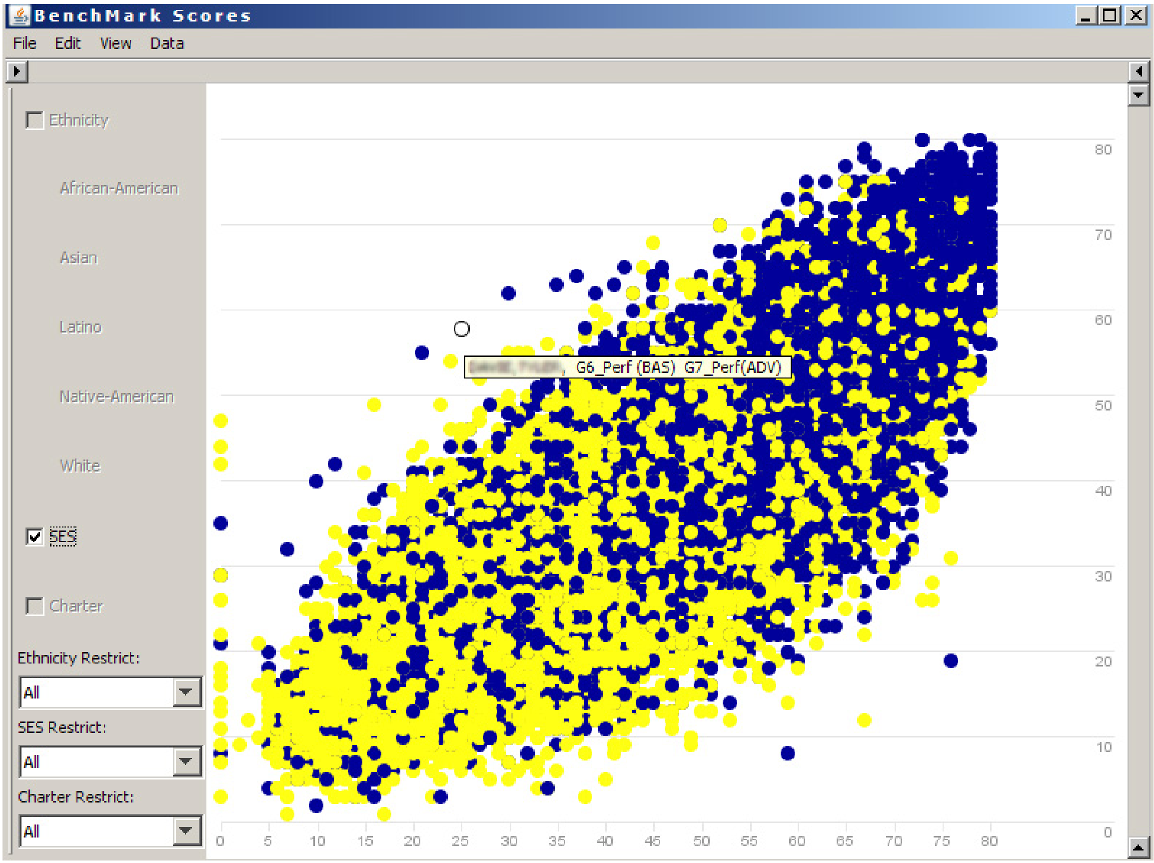


Figure 5: Modified prefuse scatter plot visualization which incorporates brushing for the encoding of particular attributes as well as mouse-hover for the discovery of individual student names. The actual student named in this example has been blurred to protect the privacy of the student.

This modified prefuse application was demonstrated to personnel within ADE, and a grant was awarded for the development of this work into the hive application. This development was done by Enspire Learning in Austin Texas, which had previously built another online application for ADE. (http://enspire.com/) Once the application was demonstrated to educators outside ADE, two more important elements were suggested. The first was that lines be added to the visualization, demonstrating the different levels of proficiency for Below Basic, Basic, Proficient, and Advanced. Since multiple students could have the same score on both the X and Y axis, it was also suggested that there be some way that plots of multiple scores be differentiated from plots that represent only a single score. This latter requested was achieved through application of the information mural algorithm. (Jerding & Stasko, 1996)

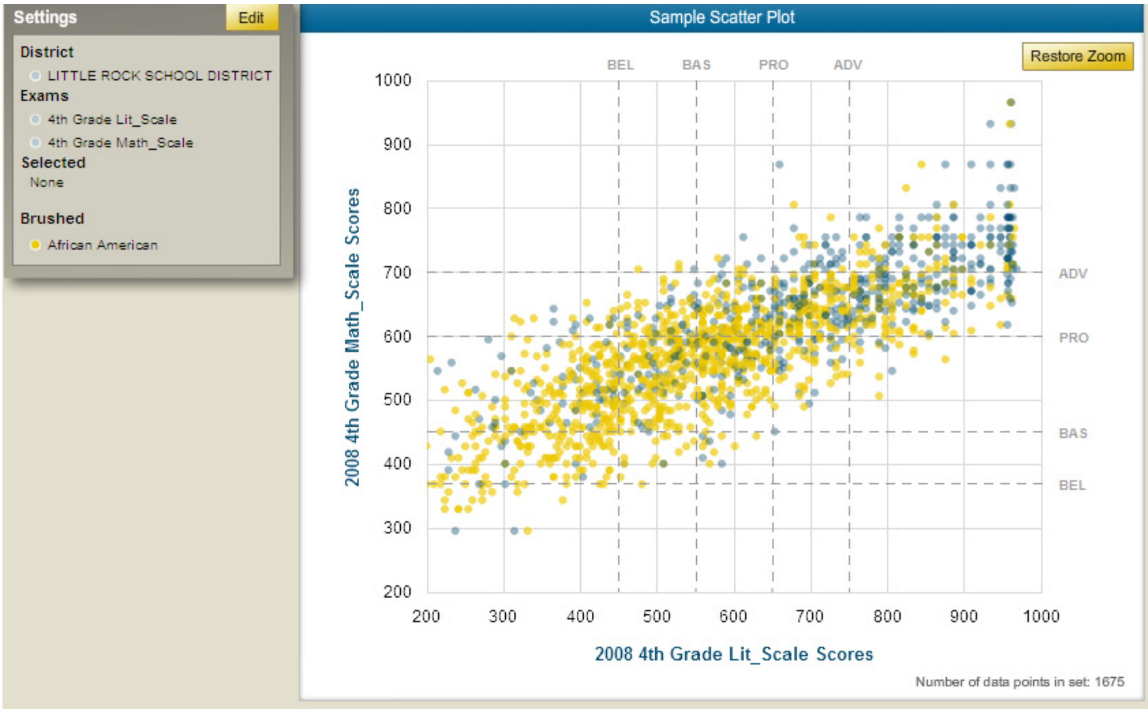


Figure 6: Scatter plot visualization of 2008 fourth grade literacy scale scores and fourth grade mathematics scales scores for the Little Rock school district. This visualization demonstrates the addition of lines denoting test proficiency as well as the application of the Information Mural algorithm which gives a particular plot a deeper hue if more than one student is represented by that plot. The mouse-hover application allows all student names for such plots to be displayed if the user is authorized.

The final suggested modification to the original prefuse scatter plot visualization was to include the ability to incorporate student growth for any particular tested subject. Although proficiency on any given examination is a product of the student's scale score for that assessment, it is important to realize that a scale score for any particular grade and subject combination is a product of not just what the student learned that particular year, it is also a reflection of the learning that student had in previous years. For example, a student's score in mathematics for sixth grade is also a product of what that student learned in the fifth grade, fourth grade, etc. Even if a student reaches proficiency for a particular subject and grade, if they also had low growth in that area, it is possible that they might slip in proficiency the next year since they did not receive expected growth.

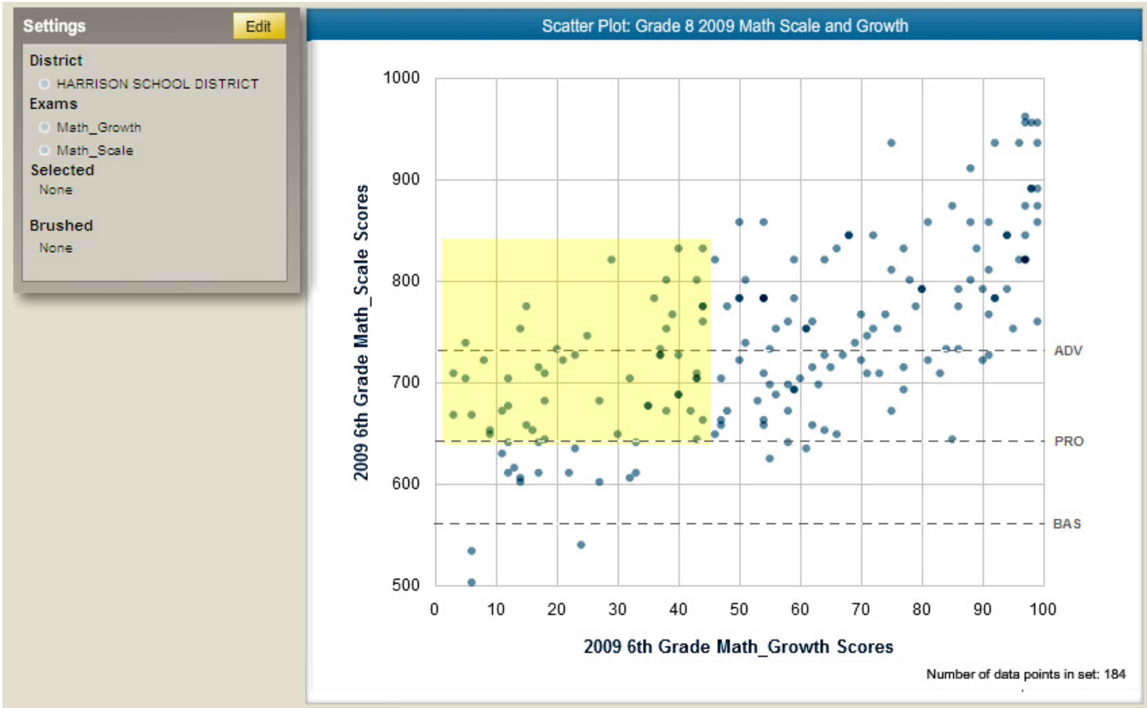


Figure 7: Scatter plot visualization of individual growth and scale scores for 2009 eighth grade mathematics in the Harrison school district. The highlighted box represents students that were proficient but which did not have average growth for that particular year and subject. Such students are at risk for slipping in proficiency since they did not reach the necessary growth needed to maintain proficiency. A mouse-hover over any individual plot will reveal the name and details of that particular student for authorized users.

Based on the feedback from those which the scatter plot was demonstrated, it was suggested that four additional visualizations be added to the program, a scatter plot over time, a bubble chart, a scatter bar, and a bar chart that included box plot overlays of the state distribution of scores. It was also suggested that having to create a new visualization was sometimes burdensome, and that instead it would be better if a user could take an existing visualization another user had created, and modify this for their own use. A user could take an existing visualization for say fourth grade mathematics and modify it so it showed fourth grade literacy or fifth grade mathematics, or take an existing visualization for one district and simply modify it by changing it to the new user's district and saving this as a new visualization.

Figure 8: Different visualizations available for users from the “Create” page. A user can instead choose to take an existing visualization and modify it for their own use without going to the Create page.

The social networking aspect of prefuse was modified to encourage threaded discussions. Users can also embed a link to another visualization within a threaded discussion. A search function was added to allow users to search for visualizations of a particular subject or grade. Besides replying to a visualization, a user can also report a visualization as inappropriate, which sends an email to the administrator. This social networking piece led to the naming of the system as "hive," as in the sense of hive computing. However, in this application the hive is not composed of a network of machines but rather distributed human analysis, educators using the system to derive meaning for their own local data which makes explicit knowledge as it is discovered, helping others understand their own data.

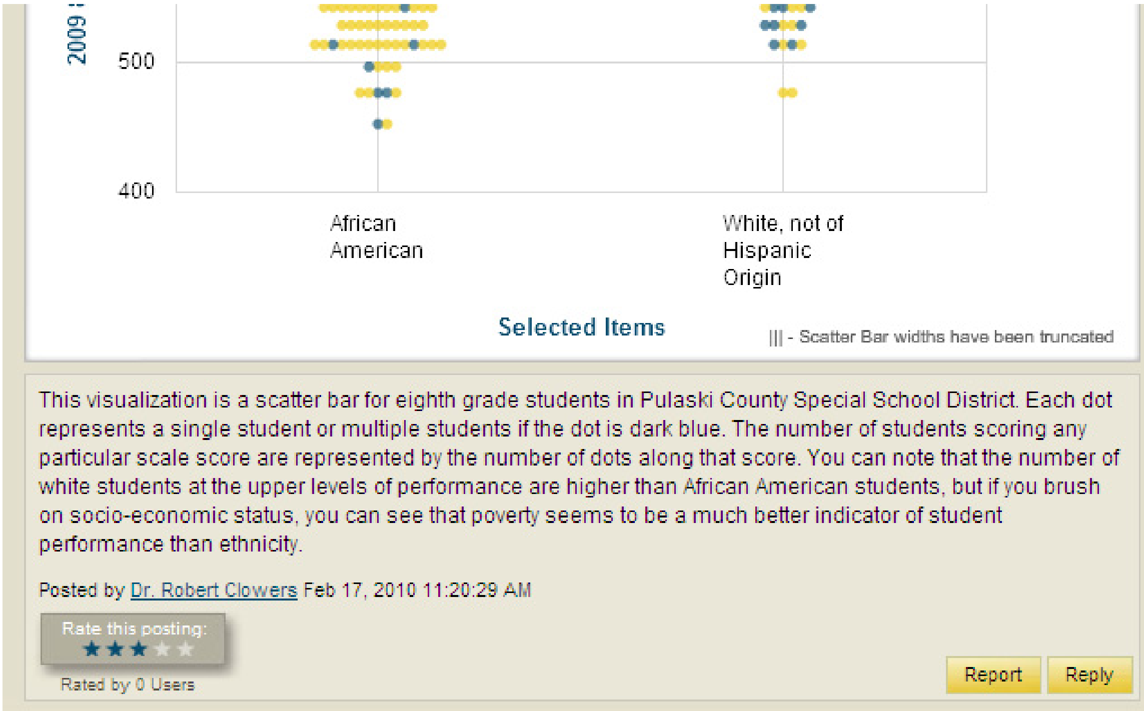


Figure 9: Visualization with a description. Note the buttons where a user can reply to this post or report it as inappropriate.

One challenge users found with hive as a tool for using the data analysis framework was that hive does not store aggregates. It was decided that for flexibility, any visualizations of aggregates would use individual student data and the calculations would be done on the client side. This solved the problem of having to continually maintain a table of aggregate values, but it also meant that many visualizations needed over 70,000 individual student scores before the visualization could be rendered. This did have another advantage in that once the data set was acquired on the client's machine, other visualizations that required individual student data could be quickly rendered with the same data set. However, in some situations, especially in schools with limited bandwidth, the time needed to create a visualization with new data was seen as a hindrance.

It was decided that instead of modifying the existing hive application for aggregate values, an adjunct website would be created that just used aggregate values of district data. Since the majority of districts have only one school for certain grades, many educators could use this site to answer questions in the framework specific to both school and district. This saved users much time in answering many questions covered by the framework, but visually it was not as compelling as what was available from the main hive website.

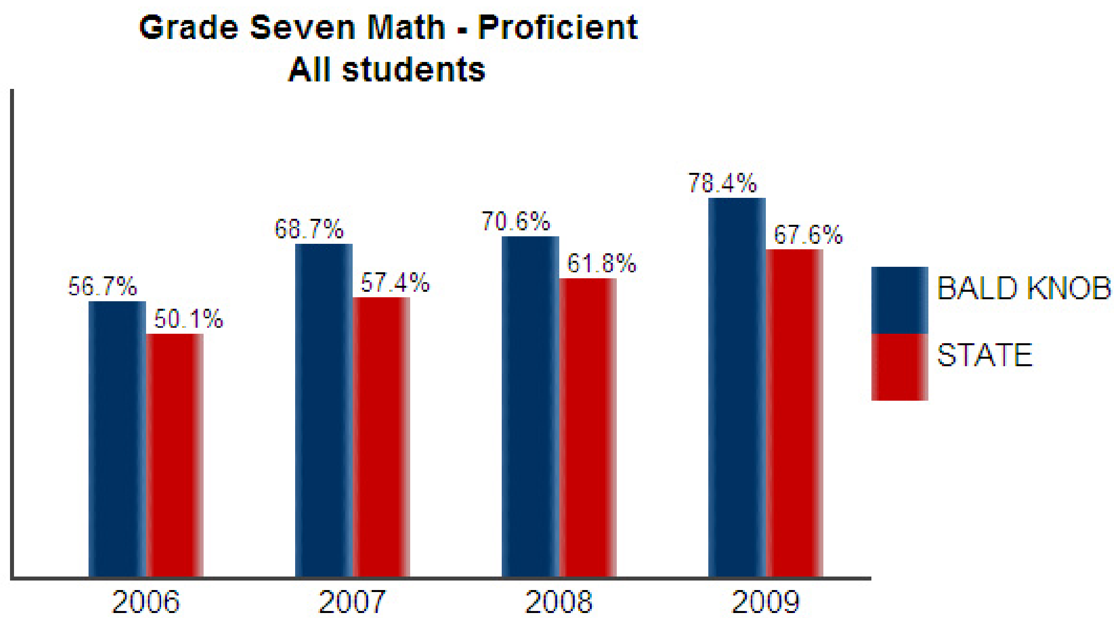


Figure 10: Adjunct visualization with aggregate values for seventh grade mathematics at the Bald Knob school district.

As users began working with this site, they again requested many changes. Given the limitations of this approach, the many different things users were requesting, and the somewhat basic visual style, it was decided that another visualization toolkit was needed for this adjunct website. After much research, FusionCharts Free was selected. (http://www.fusioncharts.com/free/) FusionCharts creates interactive Flash visualization through ActionScript, which the researchers also had experience. This provided much more flexibility for aggregate data and was much more visually compelling than the previous site. It also allowed details of the data to be discovered with a mouse-hover. This functionality was quickly incorporated into hive, and the adjunct site renamed as "QuickLooks."

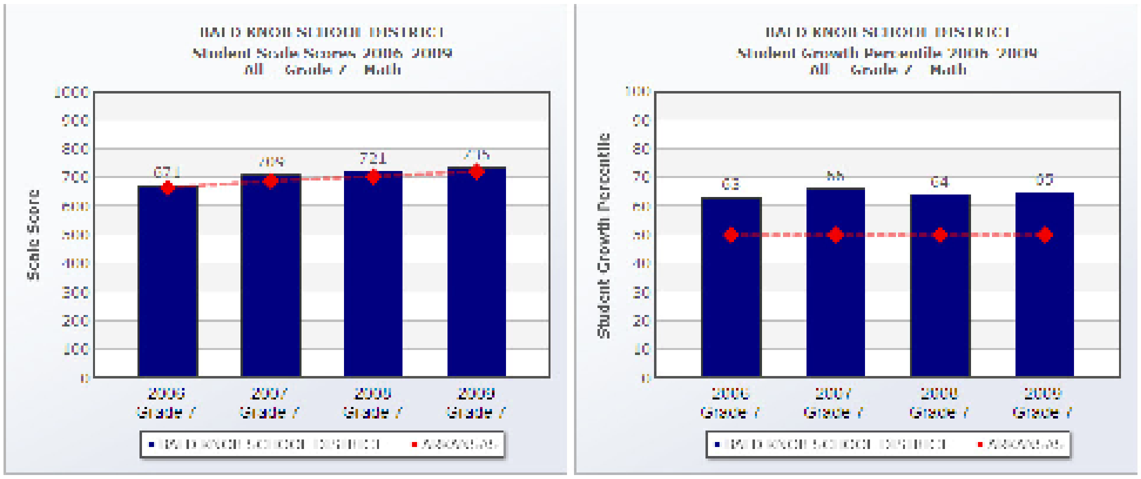


Figure 11: QuickLooks visualization for seventh grade mathematics at the Bald Knob school district. Aggregate scale scores and student growth percentiles are displayed for four years, along with the state's aggregate values represented by a line chart. A mouse-hover over either the bars or line provides details.

As with the main part of hive, adding context to a visualization is important in helping users understand its significance. Showing average student growth side by side with average scores along with state averages for both provides a great deal of context in a very efficient manner. Again, based on user input it was decided to provide much more context, such as the ability to view a cohort of students, the same group of students over time, as well as a direct comparison of district scores to another similar district. With QuickLooks, you can do both with a single visualization.

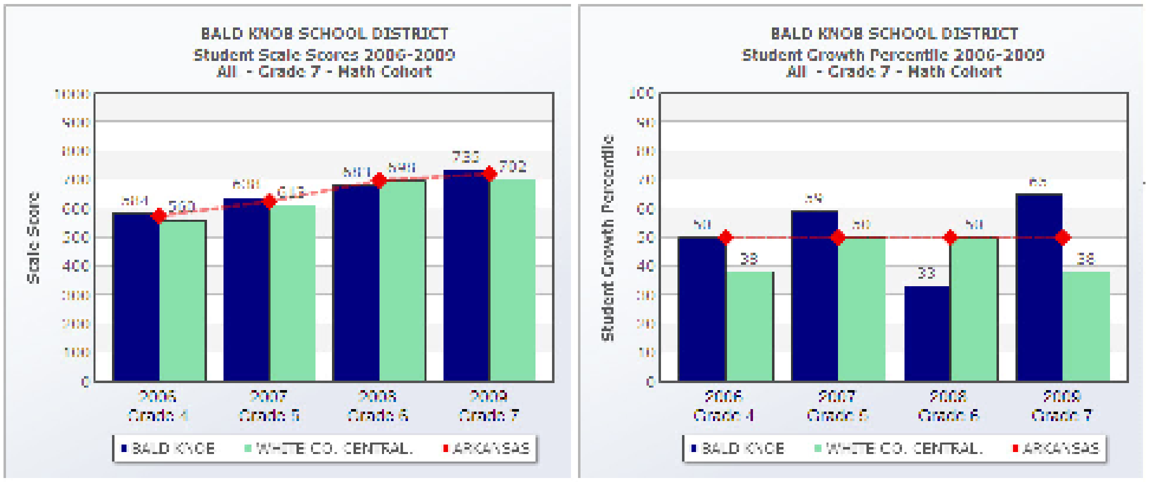


Figure 12: QuickLooks visualization of the 2009 grade seven mathematics cohort for the BaldKnob school district along with the same cohort at the White County Central school district.

While there are many other data visualization types for aggregate data within QuickLooks, including bubble charts, there was a lot of demand for including strand scores. For any given subject, there are many standards that students are supposed to learn. These learning standards are grouped under the much larger category of “strand.” For example, in algebra the strands are Language of Algebra, Solving Equations and Inequalities, Linear Functions, Non-linear Functions, and Data Interpretation and Probability. For the end of year assessment in algebra, each strand will have questions of two types, multiple choice and open response. Strand scores for both types of questions are reported, and these scores are available from within both NORMES and Triand. However, what is missing is the context, that is the ability for schools and districts to compare there strand scores with the rest of the state or the ability to compare aggregate strand scores for different levels of proficiency, all students, proficient students, and non-proficient students. Providing educators with this level of detail has had a profound impact. For example, by looking at strand scores with this level of context, it is clear that the main thing that separates proficient students from non-proficient students is that the latter group struggles with open response questions.

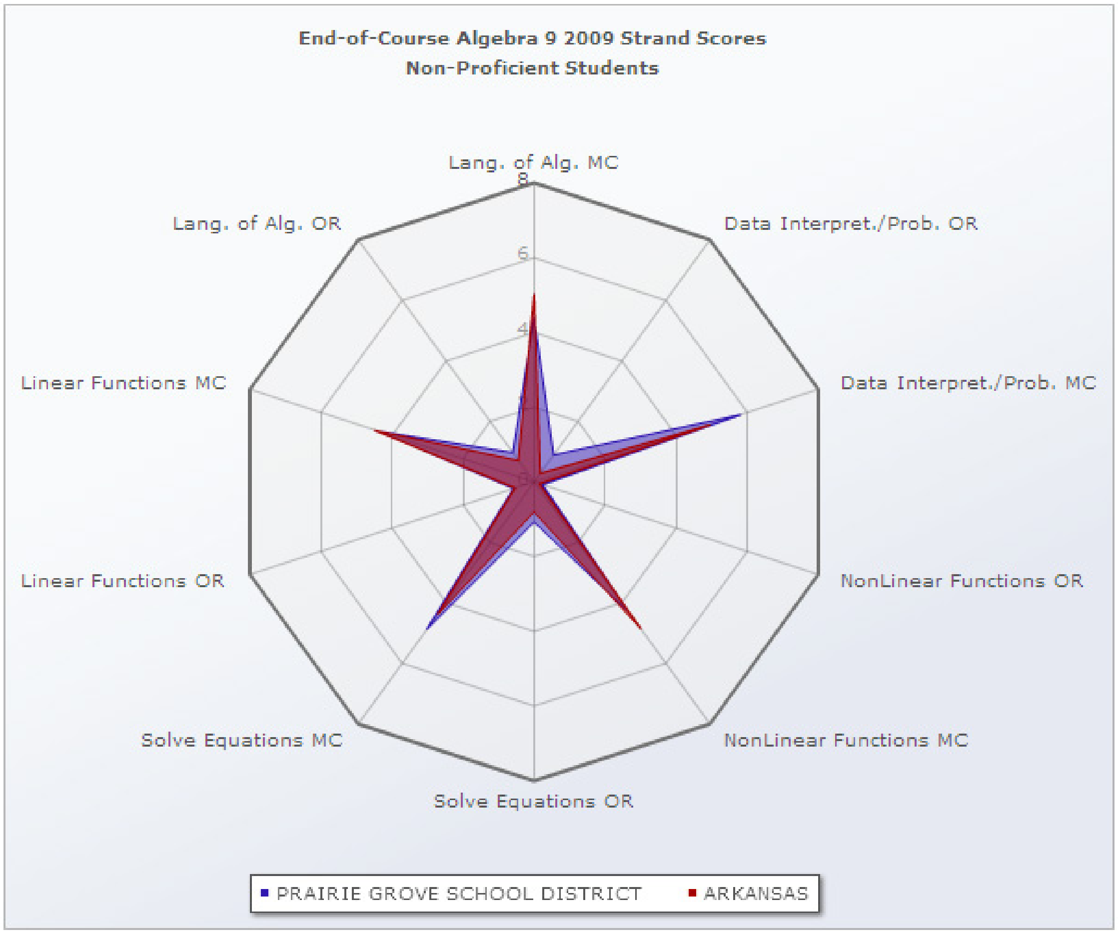


Figure 13: QuickLooks visualization of strand scores for non-proficient students in algebra at the Prairie Grove school district. Multiple choice strands are labeled "MC," and open response strands are labeled "OR." District aggregate values are in blue and state values are in red. Based on this visualization in comparison with proficient students, the primary difference between proficient students and non-proficient students are the low scores the latter group receives for open response questions.

Since both prefuse and FusionCharts Free are open-source, it was much easier to adapt the visualizations to user requests. As will be shown later, this approach of constant adaptation of hive as an IP system through iterative and incremental development has been very popular with users, who overwhelmingly gave it positive marks in comparison to other systems such as NORMES and Triand. Considering that the money allocated for NORMES is over $1,000,000 annually and Triand's cost to the state is $460,000 annually, this is not an insignificant milestone. Couple this with the fact that Arkansas' data systems were held in high regard across the nation with NORMES and Triand before the introduction of hive, is demonstrative of the fact that we took what was already considered an exemplary system and improved it significantly.

The incremental and iterative development approach of hive was done as an application of Shewhart and Deming's PDSA model. However, what was not developed during this research were good metrics for determining how much value each requested addition has made to the overall product. This is not an insignificant matter, in that it is easy to imagine a point at which requested modifications would reach a point of diminishing returns, where the addition of a new visualization or providing a new level of context did not add to the value of the overall system and instead simply served to make hive more complicated and less valuable for the consumer. We can say that hive has received countless praise from users, but we cannot ascribe any particular measure to how much a suggested modification added to the value of the overall system.

We can now return to Ge and Helfert's assertion that "interaction and information presentation, which are two important factors influencing decision making, need to be investigated as independent variables in the research of IQ effects on decision making." (Ge & Helfert, 2007, p. 13) This research and the user responses to it represent a clear improvement over system development that was not IID in nature and only delivered value at the end of development. It can also be said that each new level of context and understanding of data lead to user requests for even more detail, which means this particular IP will have to continue to adapt as its users' knowledge increases. However, what was not discovered was the point at which continued modifications to the IP led to less value for the user, that is when a change served instead to only increase the complexity of the task environment in which we are asking educators to perform.

We can demonstrate that both interaction and information presentation were improved for users with this approach to IP development. As previously noted, since 50% of data warehouse projects can be expected to fail, this alone would suggest an alternative to data warehouse development that might help avoid such failures. What this research cannot provide are metrics associated with incremental development. To date, most suggested additions to hive seemed reasonable and valuable, so they were incorporated. This has worked well for this particular IP system, but we cannot suggest that this will be the case for all system development. Much work remains in this particular area, determining the metrics associated with PDSA as it is applied to iterative and incremental development.

Chapter 5

Results

Measuring the Impact of hive

As with all aspects regarding quality, the metric of importance is "fitness for use." In regards to information products, we must query our consumers for their perceptions regarding its suitability for use. To measure the effectiveness of the data analysis framework and hive in facilitating data analysis for educators, a survey instrument was constructed to gather feedback concerning consumers' impressions of this new IP system. This survey instrument and the complete responses to it is included in Appendix B. Survey questions were grouped under two major areas. One group of questions asked respondents to compare hive with other systems they use for similar purposes. The second group asked questions concerning both the framework and hive, including questions specific to understanding if this approach would help users avoid the fundamental computational bias. Respondents were also asked if hive would facilitate communication about data, not just to other educators but parents and other stakeholders as well. Again, prior to this project Arkansas was seen as a leader in the area of education data use. Any improvements to a system already viewed as one of the best in the nation would be significant.

The data analysis framework and hive was introduced to educators in a series of short workshops held across the state from December 10, 2009 and March 15, 2010. Participation in the workshop was completely voluntary, and participation in the online survey was also completely voluntary and anonymous. There were 302 respondents to the survey. Of those, 64 identified themselves as "District Administrators," 171 identified themselves as “School Administrators,” 43 said they were “Teachers,” and 33 chose “Other.” Nine respondents choose two different roles, such as “District Administrator” and “School Administrator,” which is not uncommon in small districts where a school principal might also have a district role. It is not surprising that a majority of respondents were school administrators. NCLB accountability is focused at the school level, so school administrators have a keen interest in the education measures for their schools.

When asked what system they currently use for data analysis, 201 said “NORMES,” 70 said “Triand,” 31 said “D2SC,” 27 said “The Learning Institute,” 10 reported a variety of other systems, and 4 left this blank. Some respondents chose multiple systems, the most popular combinations being NORMES and Triand at 17 and NORMES and D2SC at 11. It is not surprising that the majority of respondents identified NORMES as their primary tool for data analysis, since the majority of respondents were administrators, and typically only administrators are given access to NORMES. For those that choose NORMES as the tool currently used for data analysis, 120 said they were school administrators and 54 said they were district administrators. The Learning Institute and D2SC are providers of interim assessments, test given every nine weeks or so to track student learning, but they are not directly associated with ADE as in the case of NORMES and Triand.

The responses to the survey were overwhelmingly positive, which is demonstrative of the fact that users found this approach beneficial. Comparing this system to what they have used in the past, 96% Strongly Agreed or Agreed that “This system provides access to data in a way that meets my needs.” This question is specific to fitness for use and specific to those that actually use the data.

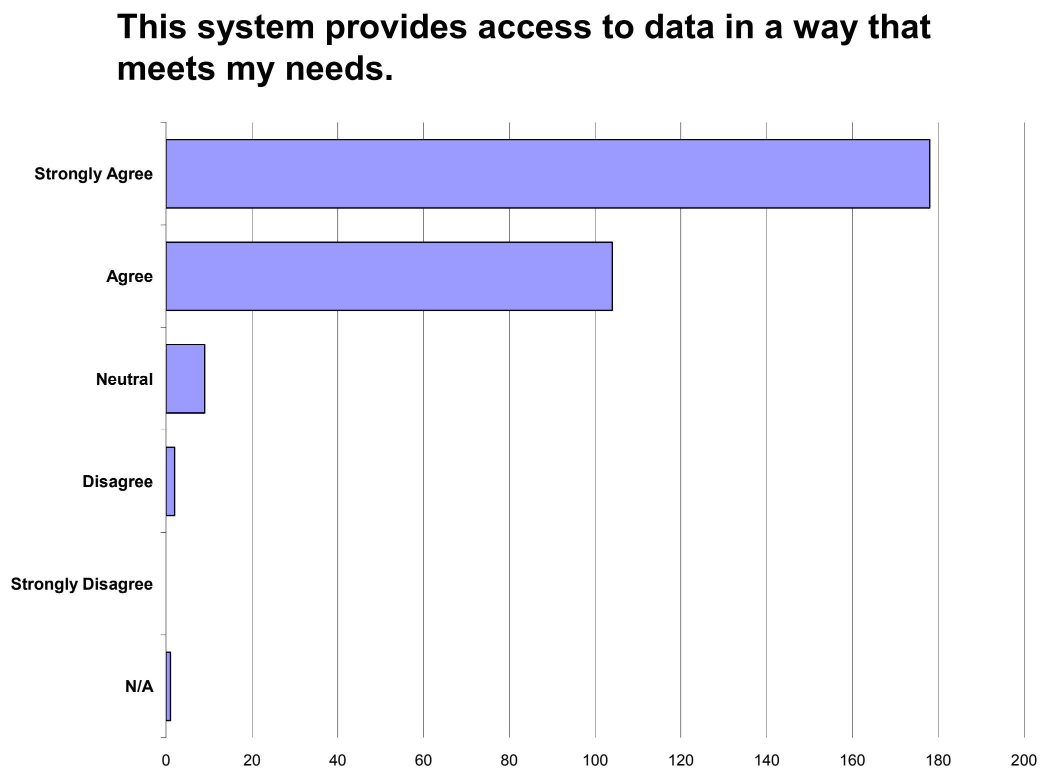
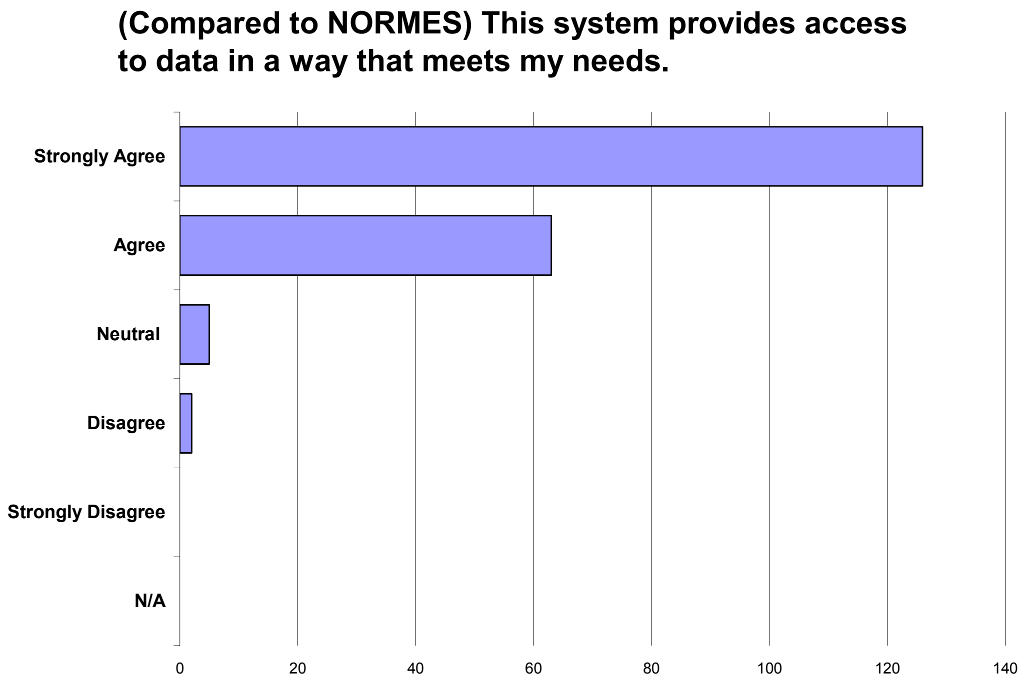


Figure 14: Summary of responses to the statement “Compared to what I have used in the past; this system provides access to data in a way that meets my needs.”

If we break out these responses by which system respondents reported they use for data analysis, when the system is being compared to either NORMES or Triand, 96% Strongly Agree or Agree that “the system provides access to data in a way that fits my needs.” There is no difference when users compare this system with NORMES or Tiand. Please note that of the 196 users that chose NORMES as their tool for data analysis and the 68 that chose Triand as their tool for data analysis, 17 of these users reported both NORMES and Triand as what they currently use for data analysis.



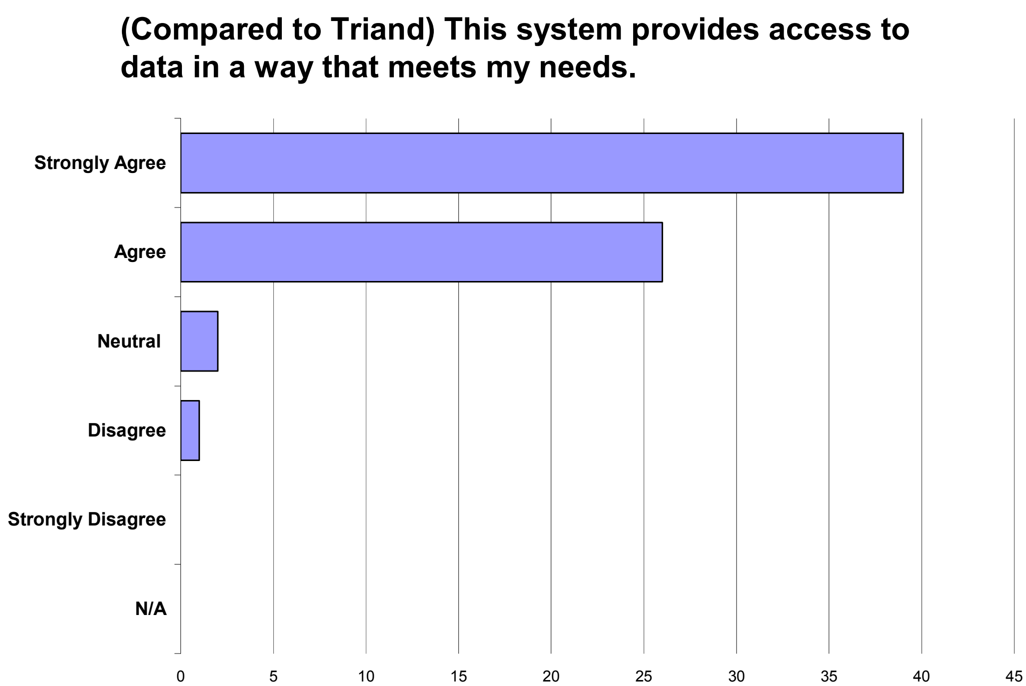


Figure 15: Summary of responses for the statement "Compared to what I have used in the past; this system provides access to data in a way that meets my needs," broken out by primary data analysis tool reported by the user. There were 17 respondents that chose both NORMES and Triand as their primary data analysis tool.

Since the combined cost for both NORMES and Triand is over $1,500,000 annually to the state, an open source tool that clearly outperforms both is significant.

The most likely reason this system is held in such high regard when compared with other systems is due to the focus on providing context for the data. This is an important step in the transformation from data to information that a user must perform internally. The incremental and iterative development of hive continues to provide feedback on different forms of context users believe would be beneficial. When asked "This system allows me to compare my school/district's performance to other schools and districts," 82% of respondents replied "Strongly Agree." No other question generated a higher percentage of Strongly Agree responses, and the percentage of those choosing "Strongly Agree" or "Agree" was also the highest at 98%. This suggests that educators do understand the importance of context, and that a primary weakness of existing IP systems may be a lack of context provided. As Tufte posits, "Data-rich designs give a context and credibility to statistical evidence. Low-information designs are suspect: what is left out, what is hidden, why are we shown so little"? (Tufte, 2006, p. 156)

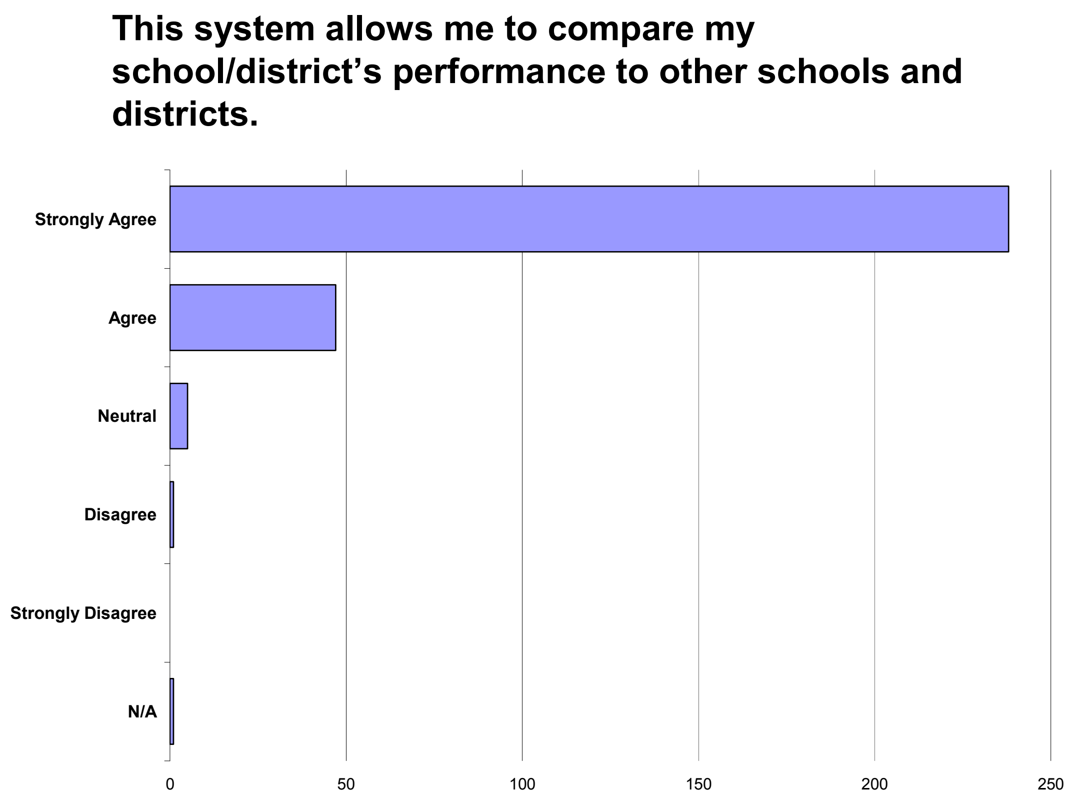


Figure 16: Summary of responses to the statement "Compared to what I have used in the past; this system allows me to compare my school/district's performance to other schools and districts.

Within this group of questions asking respondents to compare this IP with products they previously used for data analysis, the lowest number of positive responses was to the statement "This system allows stakeholders such as parents to view information about student achievement." The number of respondents that chose "Strongly Agree" was only 44% and the number choosing "Strongly Agree" or "Agree" was 75%. While these number are still relatively high, when compared to the much higher percentages for all other statements it does stand out. Concerning the entire survey, the mean for "Strongly Agree" or "Agree" for each statement was 92% and the standard deviation was 6. That responses to this particular question are a full standard deviation below the mean is indicative of how different respondents felt about this question compared to all others. We cannot assume that users simply did not realize this was a public site. This was mentioned repeatedly during the workshops, and educators were specifically encouraged to share the website's address with parents and other stakeholders. However, at this suggestion many workshop participants would seem incredulous and ask "You mean parents can see this"? While the previous responses show that educators are very positive about the context provided with hive, such as student growth percentiles and the ability to directly compare multiple districts at the same time, based on the responses to this question, it does not appear that educators are as equally comfortable knowing that parents will have the same ability.

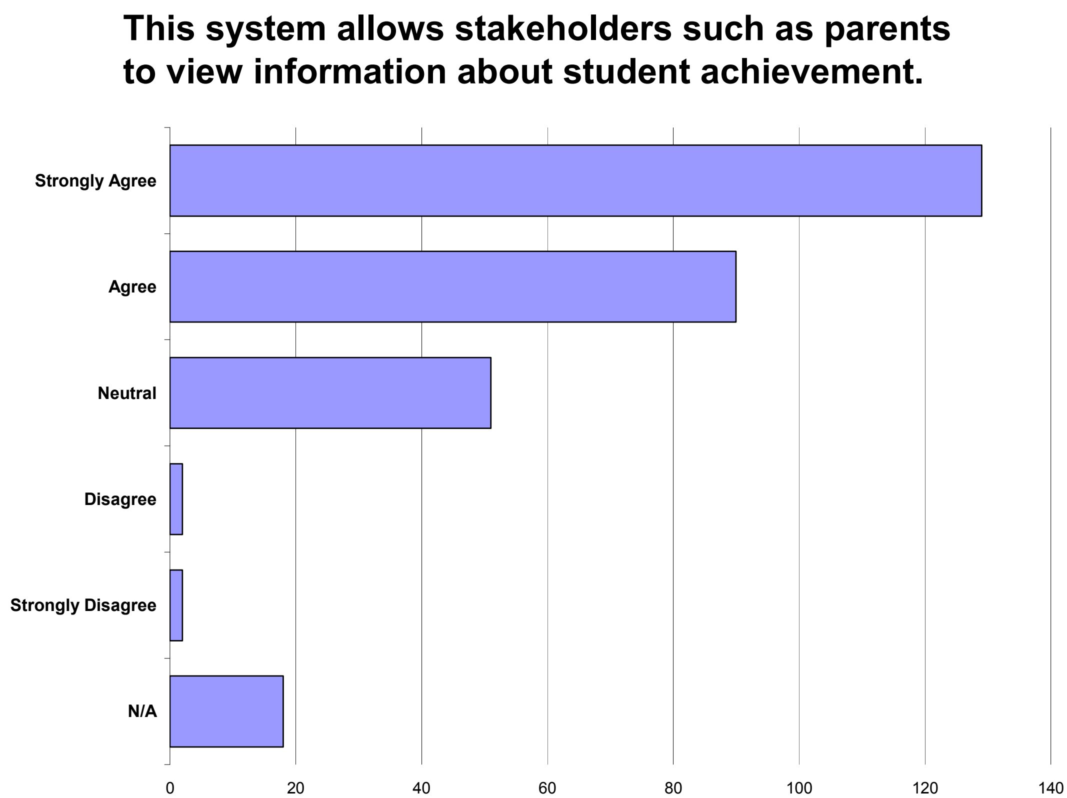


Figure 17: Summary of responses to the statement “Compared to what I have used in the past; this system allows stakeholders such as parents to view information about student achievement.

The second group of questions was specific to both the data analysis system and the visualization tools. This is a critical part of the research, since it is testing both blades of Simon’s Bounded Rationality scissors—was the combination of data analysis framework and visualization tools able to increase the computational ability of the actors and lessen the cognitive load? To answer this, the most important survey question would be “The framework and tools presented today represent a clear process for analyzing data that I did not have before.” Over 56% of respondents answered this with “Strongly Agree,” while the percentage that answered “Strongly Agree” or “Agree” was over 96%. Again, given ADE’s reputation as a national leader in data use, for users to overwhelmingly respond that the data analysis framework and tools represents something they did not have before is highly significant, made even more significant by the fact that no respondents chose “Strongly Disagree” or “Disagree” for this question.

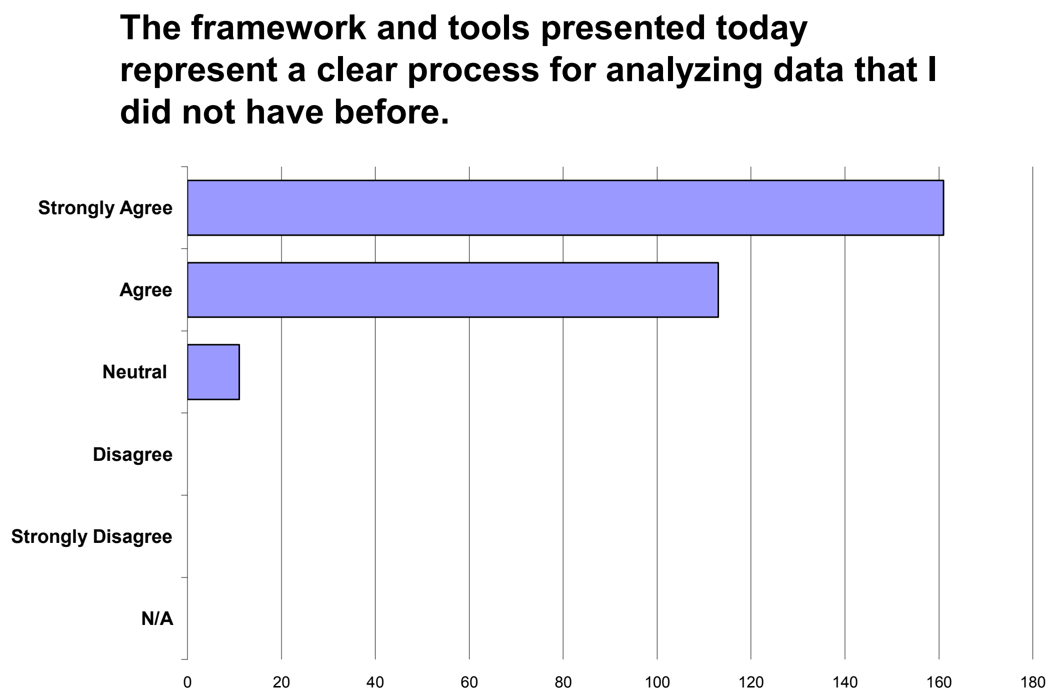
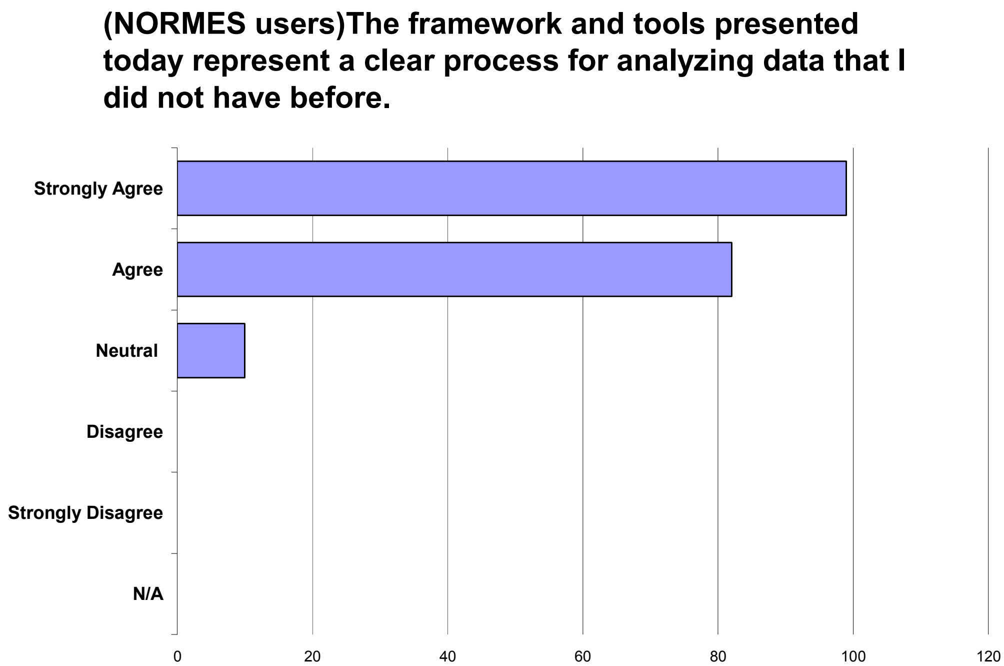


Figure 18: Summary of responses to the statement “Concerning the framework, tools, and professional development I have received today; the framework and tools presented today represent a clear process for analyzing data that I did not have before.”

Although this question did not ask respondents to specifically compare this IP with similar products they use for data analysis, it is significant to compare this approach to what users identified as their primary product for education data analysis. For those respondents that said NORMES was their primary data analysis tool, 95% “ Strongly Agree” or “Agree” that the data analysis framework and hive visualization tool represent a clear process for analyzing data that [they] did not have before, while just 90% of those that chose Triand had the same responses. As before, 17 of respondents chose both Normes and Triand as their previous tool for data analysis. Given the fact that there are much fewer respondents that chose Triand as what they use currently for data analysis than NORMES, the difference between these two different IP is not significant. What is significant is that regardless of what product users are currently using and the costs association with both NORMES and Triand, an overwhelming number of respondents found this new approach represents a clear process for data analysis they did not have before is of considerable significance.



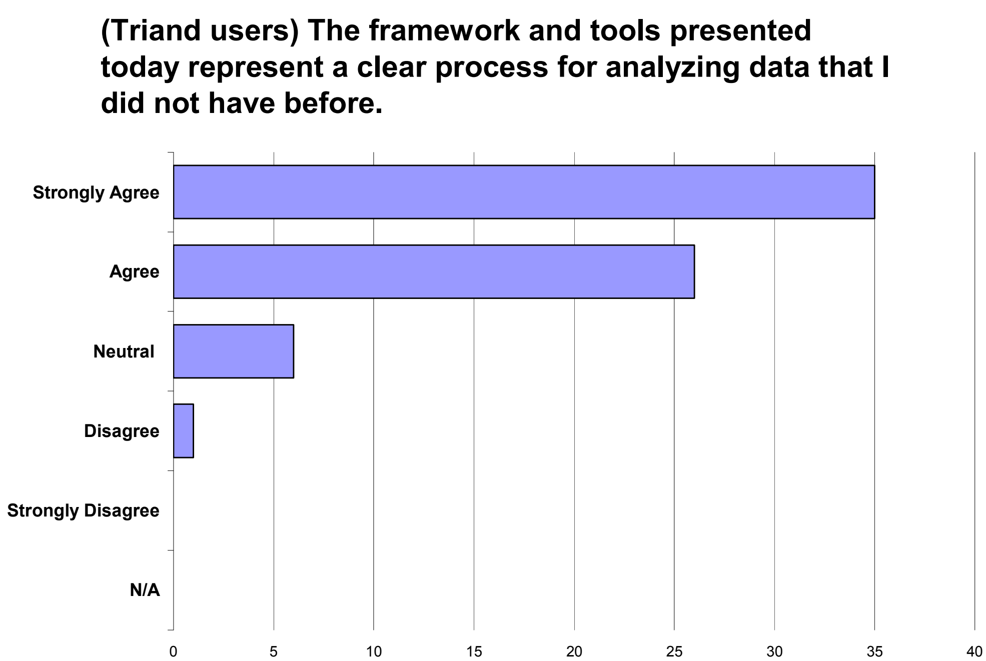


Figure 19: Summary of responses for the statement "Concerning the framework, tools, and professional development I have received today; the framework and tools presented today represent a clear process for analyzing data that I did not have before," broken out

by primary data analysis tool reported by the user. There were 17 respondents that chose both NORMES and Triand as their primary data analysis tool.

The lowest percentage of positive responses for this second set of questions was to the statement “As a result of this training, I feel comfortable showing others how to access, analyze, and use data.” Only 31% choose “Strongly Agree,” and only 79% chose “Strongly Agree” or “Agree”. This is the second lowest percentage for the entire survey. It is somewhat puzzling that users are so overwhelmingly positive about hive and process, as documented by the number of “Strongly Agree” and “Agree” responses to most other statements. However, since opportunities of training and support for data analysis for educators is somewhat rare, it is not surprising that they feel somewhat uncomfortable about analyzing data. We believe the lower numbers reflected here, in contrast with the very high numbers associated with virtually every other statement, is indicative of users gaining tacit knowledge about data analysis which they have yet to make explicit. Again, as Polanyi points out “We can know more than we can tell.” (Polanyii, 2009, p. 18)

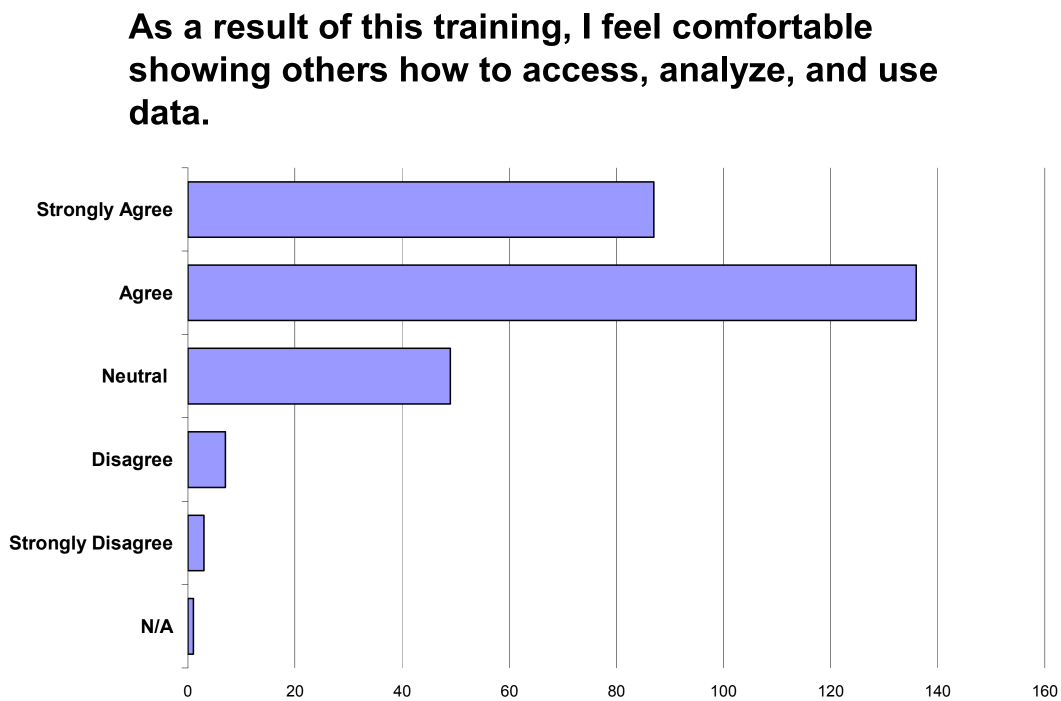


Figure 20: Summary of responses for the statement “Concerning the framework, tools, and professional development I have received today; as a result of this training, I feel comfortable showing others how to access, analyze and use data.”

Workshop participants were briefed on problems associated with the fundamental computation bias, the Jack, Anne, and George question discussed above being used as an introduction to the topic. Since helping users avoid computation bias is a major part of this research, four questions were included in the survey that were specific to measuring whether or not this IP and process would help users avoid this in the future. Respondents were very positive concerning these statements.

The first statement was “Concerning the framework, tools, and professional development I have received today; the process and tools presented today have convinced me that there is more to look at than just test scores.” This question was designed to see if user perceptions of student data had been broadened to the point where they realize they need to look at all kinds of data instead of the focus on end of year assessments so important to NCLB. This also speaks to the nature of an expanded solution set for possible actions to arise from an analysis of only a single dimension. Workshop participants where shown an example where a school decided to focus on a single subject, literacy, to raise test scores in that area, which did have the desired effect. However, workshop participants were then shown the same school's subsequent scores in mathematics, which had dropped sharply. After presenting examples of computational bias and using the new IP system, 61% of respondents chose "Strongly Agree" for this statement, and 96% of respondents chose either "Strongly Agree" or "Agree."

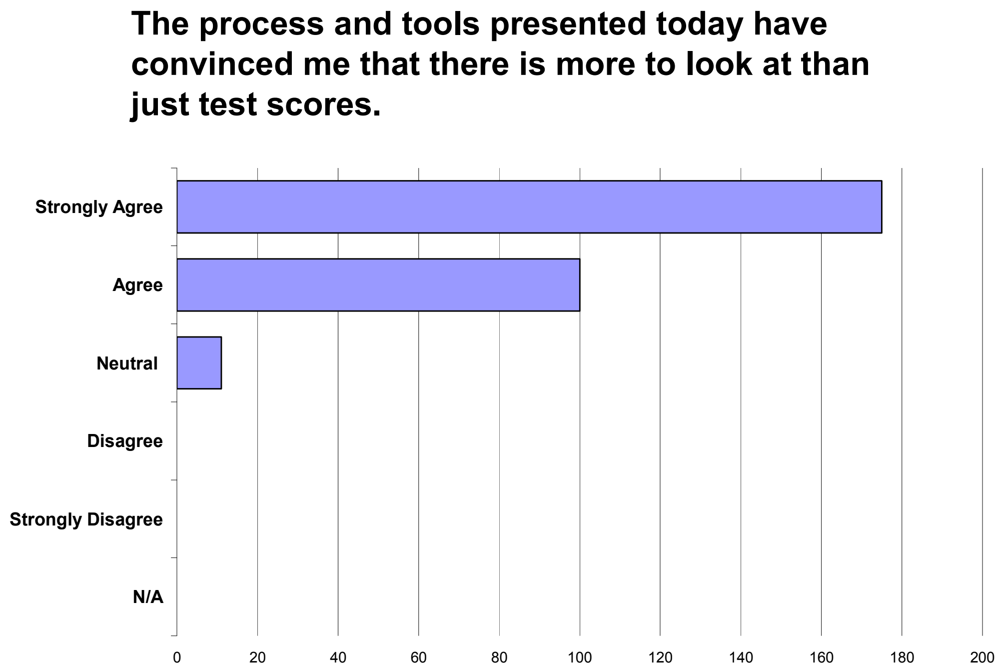


Figure 21: Summary of responses for the statement "Concerning the framework, tools, and professional development I have received today; the process and tools presented today have convinced me that there is more to look at than just test scores."

Another question asked respondents to evaluate the statement "The process and tools presented today will help me avoid unfounded assumptions when I analyze data." This question is specific to the bias of enthymematic reasoning based on unstated assumptions. One such unstated assumption is the achievement gap, for example, African American students on average have much lower scores than their white peers. However, there is nothing in the difference between scores to suggest the gap is related to ethnicity alone, and in reality is much more closely related to poverty. It was hoped that the inclusion of such things a brushing would provide educators with a different perspective on such issues. Give the responses to this statement, it would seem hive and the data analysis framework has widened educators' perceptions. 52% of respondents answered

"Strongly Agree" when asked this question, while 92% chose either "Strongly Agree" or "Agree."

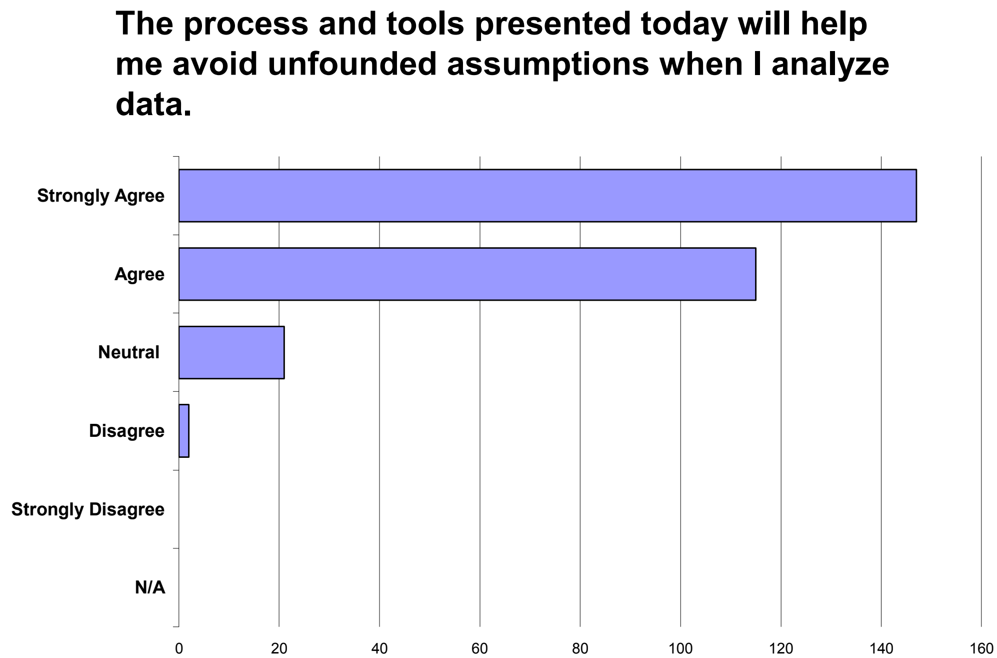


Figure 22: Summary of responses for the statement "Concerning the framework, tools, and professional development I have received today; the process and tools presented today will help me avoid unfounded assumptions when I analyze data."

Another question was included in the survey to ask respondents about problems associated with prior knowledge. This is a significant problem because of the Pygmalion effect discussed before. When looking at data, it is sometimes difficult to separate one's self from personal knowledge one may have. One of the features in hive that helps combat this is the inclusion of the mouse-hover for detail information. For example, an educator may find a point of interest in a visualization but will have to use a mouse-hover to discover which student that particular point represents. There was a great deal of discussion whenever individual names were discovered, so it was readily apparent that educators were finding examples which did not fit their previous expectations. When asked to evaluate the statement "Concerning the framework, tools, and professional development I have received today; the process and tools presented today will help me look at data while avoiding the bias of prior knowledge," 54% of respondents chose "Strongly Agree," while 94% chose either "Strongly Agree" or "Agree."

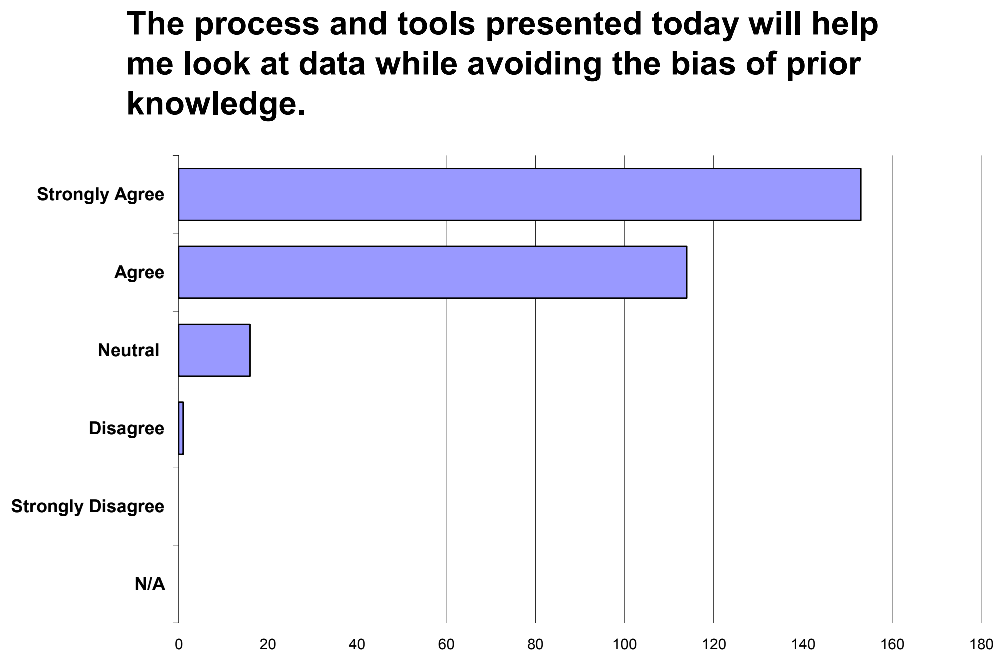


Figure 23: Summary of responses for the statement "Concerning the framework, tools, and professional development I have received today; the process and tools presented today will help me look at data while avoiding the bias of prior knowledge."

The final question specific to the fundamental computational bias is based on the tendency of users to see patterns in data that may not exist. An example of this would be Capt. Roger's belief that TN 4131 was supporting the surface vessels with which the *Vincennes* was engaged. It was believed that the many different forms of visualizations provided with hive and the ability of brushing and selection would encourage users to create multiple views of the same data, and that requiring that users do a mouse-hover to see details would encourage users be more objective about these data. When asked "Concerning the framework, tools, and professional development I have received today; the process and tools presented today will help me find patterns in the data that are different from what I expected," 61% of respondents chose "Strongly Agree" and 96% of respondents chose "Strongly Agree" or "Agree."

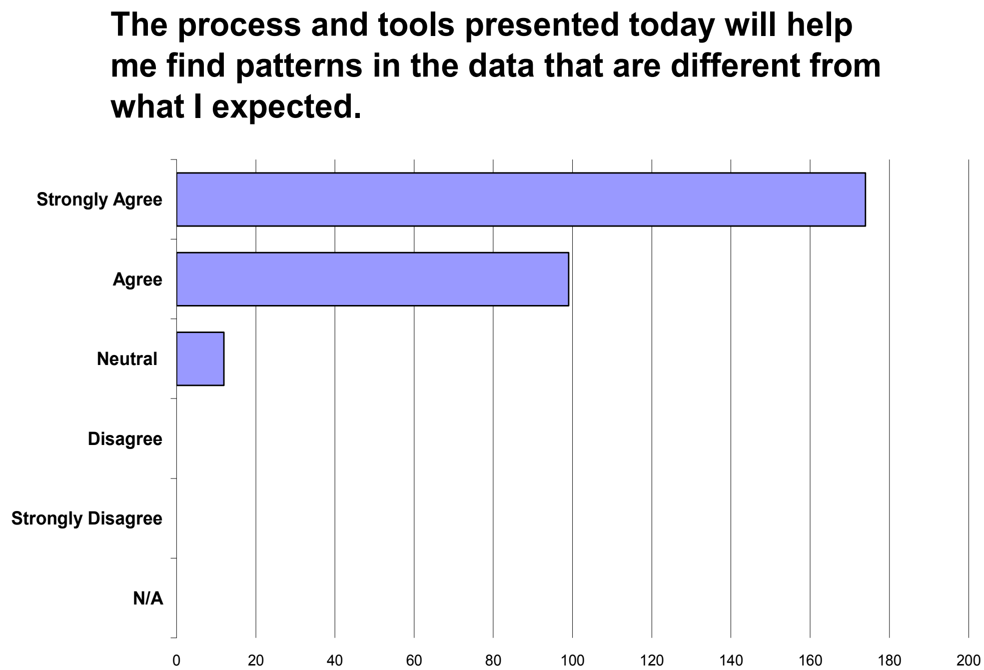


Figure 24: Summary of responses for the statement "Concerning the framework, tools, and professional development I have received today; the process and tools presented today will help me find patterns in the data that are different from what I expected."

A final section to the survey allowed users to provide open response comments. While responses to survey questions were very positive and suggest that the combination of hive and the data analysis protocol approach is significantly better than what users currently have, the actual comments that workshop participants took the time to write provide another look at the fitness of use for this data. Here are a sample of comments users provided after the workshops, and the complete set of responses is available in Appendix C:

- Would love for all data to be this simple!
- Wonderful and useful information, very user friendly. The charts and graphs and visual graphics make the data so much more understandable for teachers and administrators.
- Very user friendly and informative to help guide instruction.
- Very impressed with program. Once well-versed, this will prove an effective tool in data analysis.

- Totally excited about the possibilities for data disaggregation that this website offers.
- This was the greatest presentation of data that I have witnessed. I absolutely love the ease of moving throughout the program and the methods of data presentation. This will be a huge asset to our district. I can't wait to get back and share what I have.
- This system will allow more people to take advantage of the wealth of data at our disposal.
- This seems easy to navigate and understand. I feel that teachers will be able to easily understand and navigate through this system.
- This makes our school data less overwhelming and presents it in a manner everyone can understand and use. Great tool!
- This is the best user friendly system I have used for student data. Busy Administrators really need this.
- I think this is a great system. It will be helpful in identifying students' needs, areas of concern for our building and finding other districts that could be of help to us.
- I have been enlightened. This is the most user friendly form of data presentation I believe I have ever had the opportunity to view. This WILL be used at our school to effectively drive our instruction and training. (Gibson, 2010)

The user responses to hive and the data analysis framework have been tremendous. The creation of this IP system is clearly seen by its consumers as “fit for use” in comparison to other systems offering similar data. The iterative and incremental development of the IP allowed the developers to continually explore means of providing context that users suggested would be of benefit. We believe this approach has implications well beyond education and should be used for the development of all IP that is used to support System 2 decision process. We believe the inclusion of the data analysis framework served to increase the computational ability of the users, by providing them a well-defined problem setting and an explicit set of norms, in which to explore these data. It is believed that the social networking of hive will eventually augment these explicit norms, but the success in that area has not been as remarkable, which will be discussed in the next section.

Chapter 6

Conclusions

Research Implications

Given the user responses to the system, it is safe to say that this project was a runaway success. Since ultimately it is users that determine fitness for use, this is a significant achievement. This project has garnered interest from across the nation as well, especially since it was focused on the actual use of educational data and how to improve data driven decision making by educators. As mentioned before, using student growth as another dimension for district, school, teacher, and student evaluation has recently been embraced by the U.S. Department of Education, so ADE is now in a strong position going forward in regards to this because of this research. As an outgrowth of this project, Arkansas is the first state in the nation to tie both student scores and student growth to individual teachers, which has also prompted new research activities. However, many research questions remain unanswered.

Since the start of this research, there have been 656 users register in hive. Registration is not necessary for hive. Its adjunct site, QuickLooks, does not have registration of any kind. Registration is only necessary for users to post their own visualizations and to receive authorization to see individual student names or aggregate data with fewer than ten students represented. QuickLooks is currently averaging over 20,000 page views a month. We do not have similar statistics for the main site of hive, but we plan on building this functionality for the future. There are currently 422 individual visualizations on hive, but we do not know how many times each visualization may have been viewed by others or how many visualizations may have been created that were not saved. Again, it is how consumers actually use the system for analysis and decisions which is the primary focus, not the number of page-views it might generate.

The low number of saved posts and the fact that very few posts actually spawned threaded discussions is the most disappointing aspect of this research. We still believe that capturing discovered knowledge and making it explicit is an important part in the development of IP systems that will serve to help increase and manage organizational knowledge. We are currently developing the ability for a user to belong to a "group," such as that devoted to a particular grade, subject, or district, so that when a new visualization is generated for that area, everyone in the group will be notified and encouraged to visit. We have also not completed work on automating registration and authorization in hive. This will most likely encourage more users to register and be active, since someone within their district can authorize them instead of having to wait until a workshop is provided in their area. We expect that to quickly expand adoption of hive and the data analysis protocol, which to date have been primarily a research tool.

There are many possible explanations for the relatively low number of saved posts in hive. If we refer back to the survey question where respondents were asked if they would now be comfortable showing others "how to access, analyze and use data," we know that this is the one statement with which respondents were the least agreeable. Considering the relatively high amount of agreement on all other questions, this suggests that users are now comfortable exploring data on their own, but many may still have a reluctance sharing what they have found with others. They may need to have much more practice before they are able to make explicit what they are now able to finally comprehend at a tacit level. One somewhat common comment in the open response section of the survey was that the "workshop was too short." Guided practice, in the form of threaded discussions would serve help users gain more explicit knowledge, so expanding the use of the social networking aspects of hive remains a priority. Work has also begun on a set of online modules to be included on the AETN IDEAS portal.

Another explanation might be that this particular group, educators, is not as comfortable with something relatively new as social networking. The general population of educators tends to be much older than those normally affiliated with social networking, and educator comfort in using technology in general is also relatively low. We have already been asked by colleges that provide teacher preparation to show their students hive and the data analysis protocol, so introducing these tools to a relatively younger audience may help to initiate the social networking aspects of the tool. A core group of early adopters in this area could do a lot to spur mainstream acceptance.

Another possible explanation for the low number of saved posts could be that at the time hive was made available to the general population of educators, December 2009 to March 2010, their ability to act upon these data had already past. End of year testing for students begins in April, and it will be the results of these assessments which drive the planning districts and schools will do for the next academic year. At every workshop, participants asked when the 2010 data will be available in hive. We believe hive will be the first place educators go to view their 2010 test data once they are made available. This is also evidenced by the number of workshops for hive and the data analysis framework that have been scheduled for this summer, specific to data analysis for administrators. The plan is to have these administrators build a series of visualizations for their 2010 assessment data during these workshops, and then teachers and other administrators can do a simple search for them to see them as well and respond.

As an area of research, social networking is very young. There is no general agreement on metrics to best measure its influence and effectiveness. This is some of the first research that attempted to combine social networking with an actual production IP system, although Huang, Lee, & Wang cite many example at IBM that predate the expression "social networking" but which still exhibit some features that have since been brought under that term. (Huang et al., 1999) Based on their research and other research in knowledge management, we believe social networking to be a logical application of knowledge management moving forward. We will continue to study research in this area to determine the best metrics, and we firmly believe that the full potential of this research will not be realized until users are routinely engaging in threaded discussions about the meaning of data.

Concerning the iterative and incremental development of both hive and the data analysis framework, we believe this was critical to the project's acceptance and success, since this was the process by which proper context for visualizations was determined and the data analysis framework was modified as well. What was not discovered in this research were clear metrics to associate the PDSA cycle as it relates to iterative and incremental development. We do believe this approach of IID for information products has applications well beyond its use for education data, but we have not yet determined the exact measures for deciding how a feature might add value to an IP system as opposed to where instead it might serve only to increase the cognitive load on the user.

This is obviously an important area of research for IQ, and we will continue to explore this link in future research.

In regards to measuring the impact of the framework and hive in helping users avoid the computational bias, we believe we have had significant success in that area as evidenced by user response to those specific questions. However, it easy to assume a scenario in which a user can still be drawn into irrationality, even given the existence of good data inputs with proper context. Perhaps one area of future research in this area would be to mimic similar research well established in cognitive psychology. We might provide users with a particular set of values and ask them to arrive at a decision based on these values to measure their compliance to norms. This does represent a possible area of research, but unfortunately it would be most likely specific to the domain of information in which the scenario was based.

Information quality is an important area of research because of the central role information and technology plays in modern life. To ensure continued quality of our products requires a vigorous examination of its actual use, along with the already well-accepted understandings about the importance of quality for data inputs, the latter of which has already been well established through existing IQ research. For IP systems used in System 2 decision support, improvement for such products needs to occur along the two dimensions Simon identified—increasing the computational ability of the actors and lessening the cognitive load as they use our systems. Another important component is the ability to capture tacit knowledge and make it explicit, which in turn both increases the computational ability of decision makers and lessens their cognitive load by helping them better define the problem setting. As such, the IP system itself will be in a permanent state of beta, as it adapts to the increasing knowledge and requirements of its users.

References

Amarel, S. (1966, April). On the mechanization of creative processes. *Spectrum, IEEE*, *3*(4), 112-114.

Anderson, C. (2008, July). The end of theory: The data deluge makes the scientific method obsolete. *Wired*, *16*(7), 108-109.

Arkansas Education Television Network. (n.d.). *Arkansas ideas: Internet delivered education for Arkansas schools*. Retrieved March 13, 2010, from http://ideas.aetn.org/

Arkansas department of education: Arkansas growth model proposal. (2006). Retrieved December 3, 2009, from http://www.ed.gov/admins/lead/account/growthmodel /ar/argmp.doc

Au, W. (2007, Jun. - Jul.). High-stakes testing and curricular control: A qualitative metasynthesis. *Educational Researcher*, *36*(5), 258-267.

Bernhardt, V. L., & Geise, B. J. (March 15, 2009). *Data, Data Everywhere: ASCD Annual Conference*. Retrieved May 14, 2009, from http://eff.csuchico.edu/Downloads/ASCD%202009/ASCD%20Notes%20Mar09.pdf

Boudett, K. P., City, E. A., & Mumane, R. J. (in press). The “Data Wise” Improvement Process: Eight steps for using test data to improve teaching and learning. *Harvard Education Letter*. Retrieved May 14, 2009, from http://www.hepg.org/document/1/

Chaiken, S., & Trope, Y. (Eds.). (1999). *Dual-process theories in social psychology* (1 ed.). New York: The Guilford Press.

Cooke, N. J., & Durso, F. (2007). *Stories of modern mechnology mailures and mognitive engineering successes*. Boca Raton, Florida: CRC.

Data Quality Campaign. (2009). *The next step: Using logitudinal data systems to improve student success* (White Paper). Retrieved from http://www.dataqualitycampaign.org/files/NextStep.pdf: http://www.dataqualitycampaign.org/

Data Quality Campaign. (n.d.). *Results: Essential elements*. Retrieved February 3, 2009, from http://www.dataqualitycampaign.org/survey/compare/elements

Data Quality Campaign. (n.d.). *Ten state actions to ensure effective data use*. Retrieved January 12, 2010, from http://www.dataqualitycampaign.org/

Davenport, T. H., & Prusak, L. (1998). *Working knowledge: How organizations manage what theyknow* (1 ed.). Boston, MA: Harvard Business Press.

Department of Defense. (1988, July 28). *Formal investigation into the circumstances surrounding the downing of Iran Air Flight 655 on 3 July 1988* (Department of Defense). Washington D.C.: Department of Defense.

Diemers, D. (1999). *On the social dimension of information quality and knowledge*. Manuscript submitted for publication. Retrieved from http://www.diemers.net/sub/doc/pdf/diemers_informationqualityknowledge.pdf

English, L. (1999). *Improving data warehouse and business information quality*. New York: Wiley.

English, L. (September 2003). *Total information quality magagement--A complete methodology for IQ management*. Retrieved April 4, 2008, from http://www.information-management.com/issues/20030901/7320-1.html

Evans, D. (). *The Vincennes: A Case Study* [HTML]. Retrieved from The Internet Archive: http://web.archive.org/web/20070211035229/pobox.upenn.edu/~nrotc/ns302/20note.html

Expected utility hypothesis. (n.d.). Retrieved November 27, 2009, from http://en.wikipedia.org/wiki/Expected_utility

Filippo Pacini. (2009). In *Wikipedia*. Retrieved January 10, 2010, from http://en.wikipedia.org/wiki/Filippo_Pacini

Fisher, C. W., & Kingma, B. R. (2001). Criticality of data quality as exemplified in two disasters. . *Information and Management*, *39*(), 109-116.

Fisher, C., Lauria, E., Chengalur-Smith, S., & Wang, R. (2006). *Introduction to Information Quality* (1 ed.). Cambridge, MA: M.I.T. Information Quality Program.

Gackowski, Z. J. (2006). *Quality of informing: Credibility provisional model of functional dependencies*. Manuscript submitted for publication. Retrieved from http://informingscience.org/proceedings/InSITE2006/ProcGack194.pdf

Gartner Says More Than 50 Percent of Data Warehouse Projects Will Have Limited Acceptance or Will Be Failures Through 2007. (2005). Retrieved March 27, 2010, from http://www.gartner.com/it/page.jsp?id=492112

Gawanda, A. (2009). *The checklist manifest: How to get things right*. New York: Metropolitan Books.

Ge, M., & Helfert, M. (2007). *Develop a research agenda: A review of information quality research*. Retrieved from http://student.dcu.ie/~mouzhig2/research/ICIQ07.pdf

Gettier, E. L. (1963, June). Is justified true belief knowledge?. *Analysis*, *23*(6), 121-123. Retrieved from http://www.jstor.org/stable/3326922

Gibson, N. (2010). [Data Analysis Workshop]. Unpublished raw data.

Gibson, N. (2010). [Survey of Workshop Participants]. Unpublished raw data.

Gibson, N., & Decker, C. (2006). [Users' assessment of ADE data quality]. Unpublished raw data.

Gigerenzer, G. (2005, Spring). I think, therefore I err. *Social Research*, *72*, 195-218. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.83.5006&rep=rep1&type=pdf

Gigerenzer, G. (2008, January). Why heuristics work. *Perspectives on Psychological Science* , *3*(1), 20-29. Retrieved from http://www.unil.ch/webdav/site/determinismes/users/cclavien/public/Gigerenzer_WhyHeuristicsWork.pdf

Gladwell, M. (2007). *Blink: The power of thinking without thinking*. New York: Back Bay Books.

Guide to United States Department of Education growth model pilot program. (2009). Retrieved December 3, 2009, from http://www.ccsso.org/content/pdfs/CCSSO%20ASR%2009%20Guide%20to%20Growth%20Pilot%20projects.pdf

Heer, J., Viegas, F. B., & Wattenberg, M. (2007). Voyagers and voyeurs: Supporting asynchronous collaborative information visualization. In (Ed.), *Proceedings of ACM CHI 2007 Conference on Human Factors in Computing Systems* (pp. 1029-1038). San Jose, CA: Association for Computing Machinery. Retrieved from http://vis.berkeley.edu/papers/sense.us/2007-sense.us-CHI.pdf

Henkoff, R. (1993, June 28). The hot new seal of quality. *Fortune*, *127*(13), 116-120.

Heritage, M. (2009). *Cover sheet to accmpany materials sent to review panel for sample projects*. Los Angeles, CA: Assessment and Accountability Comprehensive Center.

Huang, K., Lee, Y. W., & Wang, R. Y. (1999). *Quality information and knowledge*. Upper Saddle River, NJ: Prentice Hall.

Ivanov, K. (1972). *Quality-Control of information: On the concept of accuracy of information in data-banks and in management information systems* (Doctoral dissertation, University of Stockholm Sweden). Retrieved from http://www.informatik.umu.se/~kivanov/diss-avh.html

Jerding, D. F., & Stasko, J. T. (1996). *The Information Mural*. Retrieved from http://www.cc.gatech.edu/gvu/softviz/infoviz/information_mural.html

Johnson, S. (2006). *The ghost map : The story of London's most terrifying epidemic and how it changed science, cities, and the modern world* (1st ed.). London: Riverhead Books.

Klein, B. D., Goodhue, D. L., & Davis, G. B. (1997, June). Can humans detect errors in data? Impact of base rates, incentives, and goals. *MIS Quarterly*, *21*, 169-194. Retrieved from http://www.misq.org

Kuhn, T. (1996). *The structure of scientific revolutions* (3 ed.). Chicago: University of Chicago Press.

Larman, C., & Basili, V. R. (2003, June). Iterative and incremental development: A brief history. *Computer*, *36*(6), 47-56.

Mandinach, E. B., Honey, M., & Light, D. (April 9, 2006). *A Theoretical Framework for Data-Driven Decision Making*. Unpublished manuscript. Retrieved October 6, 2008, from http://cct.edc.org/admin/publications/speeches/DataFrame_AERA06.pdf

Marsh, J. A., Pane, J. F., & Hamilton, L. S. (2006). *Making Sense of Data-Driven Decision Makingin Education: Evidence from Recent RAND Research*. Retrieved May 16, 2009, from http://www.rand.org/pubs/occasional_papers/2006/RAND_OP170.pdf

National Center for Education Statistics: Institute of Education Sciences (n.d.). *Statewide Longitudinal Data Systems Grant Program: Program Overview* . Retrieved May 24, 2009, from http://nces.ed.gov/programs/slds/

National Center for Educational Statistics. (2006, September). *Forum Guide to Decision Support Systems* (White Paper). Retrieved from http://nces.ed.gov: http://nces.ed.gov/pubsearch

Neumann, J. V., & Morgenstern, O. (1955). *Theory of games and economic behavior* (3 ed.). Princeton, NJ: Princeton University Press.

No Child Left Behind. (n.d.). Retrieved April 14, 2009, from http://ed.gov/nclb/landing.jhtml

Nonaka, I. (1994, February). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14-37. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download;jsessionid=D00FD6A7100A887E59A5630180BE38F4?doi=10.1.1.115.2590&rep=rep1&type=pdf

Orr, K. (1998, February). Data qauality and systems theory. *Communications of the ACM*, *41*(2), 66-71. Retrieved from http://portal.acm.org

Payne, J. W., & Bettman, J. R. (2007). Walking with the scarecrow: The information-processing approach to decision research. In (Ed.), *Blackwell handbook of judgment and decision making* (pp. 110-130). Hoboken, NJ: Wiley-Blackwell.

Pipino, L. L., Lee, Y. W., & Wang, R. Y. (2002). Data Qualit Assessment. *Communications of the ACM*, *45*, 211-218.

Polanyi, M. (2009). *The tacit dimension* (Reissue ed.). Chicago: University of Chicago Press.

Redman, T. (2001). *Data quality: The field guide*. Woburn, MA: Digital Press.

Redman, T. C. (1992). *Data quality: Management and technology*. New York: Bantam Books.

Rogers, W., & Rogers, S. (1992). *Storm Center: The USS Vincennes and Iran Air Flight 655*. Anapolis, Maryland: Naval Institute Pres.

Rosenthal, R., & Jacobson, L. (2003). *Pygmalion in the classroom: Teacher expectation and pupils' intellectual development*. Carmarthen, UK: Crown House Publishing.

Rufe, P. D. (2002). *Fundamentals of manufacturing* (2 ed.). Dearborn MI: Society of Manufacturing Engineers.

Schon, D. (1995). *The reflective practitioner: How professionals think in action*. Surrey, UK: Ashgate Publishing.

Seward, W. H. (1858, October 25). *On the irrepressible conflict*. New York History Net. Albany, NY.

Simon, H. (1996). *The sciences of the artificial* (3 ed.). Cambridge, MA: The MIT Press.

Simon, H. A. (1978, May). Rationality as process and as a product of thought. *The American Economic Review*, *68*(2), 1-16.

Simon, H. A. (1990). Invariants of Human Behavior. *Annual Review of Psychology*, *41*(), 1-19.

Spence, R. (2007). *Information visualization: Design for interaction* (2 ed.). Essex, UK: Pearson Education Limited.

Stanovich, K. E. (1999). *Who is Rational? Studies of Individual Differences in Reasoning*. MahWah, NJ: Lawrence Erlbaum Associates.

Stanovich, K. E. (2003). The fundamental computational biases of human cognition:Heuristics that (sometimes) impair decision making and problem solving. In J. E. Davidson & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 291-342). New York: Cambridge University Press.

Stanovich, K. E. (2009). *What Intelligence Tests Miss: The Psychology of Rational Thought*. New Haven: Yale University Press.

Statewide longitudinal data systems grant program. (n.d.). Retrieved March 22, 2009, from http://nces.ed.gov/Programs/SLDS/index.asp

Taleb, N. N. (2007). *The black swan: The impact of the highly improbable* (1 ed.). New York: Random House.

The Colorado growth model: Higher expectations for all students. (2008). Retrieved December 3, 2009, from http://www2.ed.gov/admins/lead/account/growth model/co/cogrowthproposal101508.pdf

The John Snow Archive and Research Companion. (n.d.). Retrieved March 13, 2010, from http://johnsnow.matrix.msu.edu/images/online_companion/chapter_images/fig12-5.jpg

The MIT total data quality management program. (n.d.). Retrieved April 4, 2008, from http://web.mit.edu/tdqm/www/about.shtml

The Pew Center for the States. (2010). *The trillion dollar gap: Underfunded state retirement systems and the roads to reform*. Retrieved from The Pew Center for the States website: http://downloads.pewcenteronthestates.org/The_Trillion_Dollar_Gap_final.pdf

Tufte, E. R. (2006). *The visual display of quantitative information* (2 ed.). Graphics Press: Cheshire, CT.

Tuomi, I. (1999). Data is more than knowledge: Implications of the reversed knowledge hierarchy for knowledge management and organizational memory. *Journal of Management Information Systems*, *16*(3), 107-121.

U.S. Chamber of Commerce. (2009, November). *Leaders and laggards: A state-by-state report card on educational innovation* (Report). Retrieved from U.S. Chamber of Commerce website: http://www.uschamber.com

U.S. Department of Education. (2009). *Race to the top: Executive summary*. Retrieved January 12, 2010, from http://www2.ed.gov/programs/racetothetop/executive-summary.pdf

U.S. Department of Education. (2009). *State grants under the State Fiscal Stabilization Fund: Summary of requirements*. Retrieved January 12, 2010, from http://ed.gov/programs/statestabilization/summary-requirements.doc

Von Houten, L., Mlyasaka, J., Agullard, K., & Zimmerman, J. (n.d.). *Developing an effective school plan: An activity-based guide to understanding your school and improving student outcomes*. Retrieved from http://www.wested.org/cs/we/view/rs/795

Vygotsky, L. (1986). *Thought and language* (A. Kozulin Trans.) (Revised ed.). Cambridge, MA: The MIT Press. (Original work published 1986)

Wang, R. (1998, February). A product perspective on total data quality management. *Communications of the ACM*, *41*(2), 58-65.

Wang, R. Y., & Strong, D. M. (1996, Spring). Byeond accuracy: What data quality means to data consumers. *Journal of Management Information Systems*, *12*(4), 5-34. Retrieved from http://w3.cyu.edu.tw/ccwei/PAPER/ERP/data%20quality%28JMIS%29.pdf

Ware, C. (2004). *Information visualization: Perception for design*. San Fancisco: Morgan Kaufmann Publishers.

Zweig, J. (2002, June). What fund investors really need to know. *Money Magazine*, *31*(6), 110-124. Retrieved from http://www.intelligenthedgefundinvesting.com/pubs/rb-jz.pdf

Appendix A

Data Analysis Framework and Protocol

Data Analysis Framework

Major Focus	Questions	Methodology	Interpretation
How are our students performing by subject area?	How do our students compare to other students in the district or at the state level? How do our students compare to other similar schools in the state? How do our students do in a particular subject across all grade levels within the school?	In Quick Looks, look at the aggregate numbers for you district. In hive, make a bubble chart of scale v. growth by grade and subject and begin selecting schools or districts to compare against your own. In hive, make a scatter plot of math v. literacy for your school/district to determine possible problem areas. In Quick Looks, compare growth and performance by grade level for a particular subject.	Or some grades/subjects performing substantially better or worse than others? Are other similar schools consistently out performing your own school? Is performance for some grades/subject stronger than others?
What are the trends in student performance over time?	What do trends look like by grade level? How do trends compare with the state and district? How do trends compare with other similar schools in the state Do particular cohorts perform above or below the average?	In Quick Looks, see if the trends for your district's grade levels are similar to the trends at the state level. In hive, you can create a scatter plot over time to see patterns of potential problems. In hive, create a bubble chart for each subject, with scale score on the Y axis and growth on the X axis. Use this to compare your district/school to others in the state. In Quick Looks, view performance by cohort (e.g. 4th graders 2006, compared to 5th graders 2007 and 6th graders 2008) to see if there are weak or strong areas.	Are there particular grades/subjects which are stronger or weaker? Is our district/school's performance similar to the state or other districts/schools? Is our growth and performance similar to that which is happening across the state? (Remember, the general pattern has been for students in the state to perform better each year.) Do we have cohorts which are stronger or weaker than others?

How are sub groups performing over time?	How are our sub groups currently performing compared to the district/state as well as similar districts/schools? What are the trends for our subgroups over time compared to state trends? How do the trends for our subgroups compare to other similar districts and schools over time? What about subgroups within the school?	In Quick Looks, you can view sub groups compared to state levels. If your school/district is predominately a single sub group, you should do most of your comparison to this sub group and not the overall state averages. In hive, you can use a scatter plot you made above and simply use brushing or selection on a sub group to make them stand out from the rest. In Quick Looks, see the performance by sub group over time to compare against the state averages. In hive, you can create multiple scatter plots over time and use brushing and selection to make sub groups stand out. In hive, you can make box plots to compare sub group performance to state averages. In Quick Looks, you can note your own averages for sub groups and then begin looking at other districts for comparison. In hive, you can make bubble charts by sub groups for multiple schools/districts. In hive, you can make box plots over time to determine if sub group performance is stronger or weaker than other districts/schools.	Is our performance by subgroups better or worse than the state averages for these sub groups? Are we closing the achievement gap or simply reflecting what is happening at the state level? Are our achievement gaps actually worse than what is seen at the state level? Over time, have we shown progress in closing the achievement gap are simply reflecting the same patterns at the state level? Are there similar districts/schools that are having success closing the achievement gap?
What are our strengths and weaknesses in teaching and learning?	How do our students perform by strand? What are the trends over time? How does our performance by strand compare to similar districts/schools?	In Quick Looks, view strand scores in comparison to the state averages. In hive, use the box plot strand scores visualization to view strand scores in comparison to state averages. In Quick Looks, view strand scores in	Are there particular strand scores that are particularly weak or strong compared to the state averages? Are there other districts/schools that are much stronger in areas your particular school is weak?

How are sub groups performing over time?	How are our sub groups currently performing compared to the district/state as well as similar districts/schools? What are the trends for our subgroups over time compared to state trends? How do the trends for our subgroups compare to other similar districts and schools over time? What about subgroups within the school?	In Quick Looks, you can view sub groups compared to state levels. If your school/district is predominately a single sub group, you should do most of your comparison to this sub group and not the overall state averages. In hive, you can use a scatter plot you made above and simply use brushing or selection on a sub group to make them stand out from the rest. In Quick Looks, see the performance by sub group over time to compare against the state averages. In hive, you can create multiple scatter plots over time and use brushing and selection to make sub groups stand out. In hive, you can make box plots to compare sub group performance to state averages. In Quick Looks, you can note your own averages for sub groups and then begin looking at other districts for comparison. In hive, you can make bubble charts by sub groups for multiple schools/districts. In hive, you can make box plots over time to determine if sub group performance is stronger or weaker than other districts/schools.	Is our performance by subgroups better or worse than the state averages for these sub groups? Are we closing the achievement gap or simply reflecting what is happening at the state level? Are our achievement gaps actually worse than what is seen at the state level? Over time, have we shown progress in closing the achievement gap are simply reflecting the same patterns at the state level? Are there similar districts/schools that are having success closing the achievement gap?
What are our strengths and weaknesses in teaching and learning?	How do our students perform by strand? What are the trends over time? How does our performance by strand compare to similar districts/schools?	In Quick Looks, view strand scores in comparison to the state averages. In hive, use the box plot strand scores visualization to view strand scores in comparison to state averages. In Quick Looks, view strand scores in	Are there particular strand scores that are particularly weak or strong compared to the state averages? Are there other districts/schools that are much stronger in areas your particular school is weak?

	How do our sub groups perform by strand? What are the trends over time?	comparison to other districts. In hive, use the box plot strand scores visualization to see if other schools/districts are more successful in certain strands. In hive, use the box plot strand scores visualization and select subgroups to see the comparison. In Quick Looks, you can see strand score data by sub group compared to state averages. In Quick Looks, note problem areas in the current year's data, and then look at prior years to see if this problem has persisted. In hive, create multiple box plot strand scores visualizations to see the trends over time. In Quick Looks, note problem areas in the current year's data, and then look at prior years to see if this problem has persisted.	What might you do to determine causes for this? Are there persistent areas of strength or weakness in strand scores for our sub groups? What might you do to determine a cause for this? What other data might you need to collect?
Can we evaluate teacher, program, and instructional strategies effectiveness?	What changes would you expect to see based on interventions you have initiated? Using longitudinal analysis, can you point to a specific change in program or instructional strategies that have been successful or unsuccessful? Can you determine which teachers seem stronger or weaker in certain areas over time?	If you initiated an important intervention for a particular academic year, does the longitudinal analysis provide sufficient evidence as to the efficacy of that change? If you are a small school/district and can identify a single teacher responsible for a particular grade/subject are the data you've analyzed above sufficient to identify strength or weaknesses of individual teachers? If you are a larger school/district, with multiple teachers for each grade/subject, would the ability to aggregate results at the teacher level be helpful?	If the data are sufficient to identify particular changes that have lead to either increased or decreased student achievement, what might you do to determine what additional actions need to take place? If the data are not sufficient to accurately determine the efficacy of program/instructional changes, what additional data might need to be collected to help you determine the value of changes? If you can identify individual teacher effectiveness, what might

			you do to determine what corrective actions might need to be taken for weak teachers and what might you do to identify the practices that make some teachers more effective? If you cannot identify individual teacher performance, would this be helpful and do your schools and the state need to do more to provide this?

GUIDELINES FOR WRITING DATA STATEMENTS AND SUMMARIES

Each statement should …

- Communicate a single idea about student achievement
- Present the facts objectively rather than state evaluative or explanatory comments
- Be short, clear sentences or phrases in everyday language that is easy to understand
- Be an independent statement, that is, its meaning should not be dependent on other statements
- Represent the data accurately by including relevant numerical data when needed for evidence
- Review all of the data statements and identify the most important ideas that convey the story about achievement

☐ Write a paragraph of statements summarizing the major and important findings. The statements can be in a slightly more narrative style, but still tightly based on data. Important numerical results should be included to support the points made. Avoid including personal judgments and opinions. If you find you are describing why the results occurred, or using the word "because" in your summary, you have moved to interpretation and are no longer summarizing!

Appendix B
Survey Instrument and Responses

Compared to what I have used in the past:						
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
This system provides a comprehensive picture of student achievement.	61.6% (181)	36.4% (107)	1.4% (4)	0.3% (1)	0.0% (0)	0.3% (1)
This system provides access to data in a way that meets my needs.	60.5% (178)	35.4% (104)	3.1% (9)	0.7% (2)	0.0% (0)	0.3% (1)
This system facilitates a straightforward method for analyzing data.	59.9% (175)	35.6% (104)	3.4% (10)	0.7% (2)	0.0% (0)	0.3% (1)
This system provides opportunities to collaborate with my colleagues in my district/school about data.	63.3% (186)	33.3% (98)	2.7% (8)	0.0% (0)	0.0% (0)	0.7% (2)
This system provides opportunities to collaborate with colleagues outside my school/district about data.	57.8% (170)	34.7% (102)	6.1% (18)	0.7% (2)	0.0% (0)	0.7% (2)
This system allows me to compare my school/district's performance to other schools and districts.	81.5% (238)	16.1% (47)	1.7% (5)	0.3% (1)	0.0% (0)	0.3% (1)
This system provides information in a way that I can use to help make changes to instruction or programs.	58.7% (172)	35.5% (104)	4.4% (13)	0.7% (2)	0.3% (1)	0.3% (1)
This system provides information about students that I can use to individualize instruction.	53.4% (156)	30.5% (89)	13.4% (39)	1.4% (4)	0.0% (0)	1.4% (4)
This system allows stakeholders such as parents to view information about student achievement.	44.2% (129)	30.8% (90)	17.5% (51)	0.7% (2)	0.7% (2)	6.2% (18)

Concerning the framework, tools, and professional development I have received today:						
	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree	N/A
The framework and tools presented today represent a clear process for analyzing data that I did not have before.	56.5% (161)	39.6% (113)	3.9% (11)	0.0% (0)	0.0% (0)	0.0% (0)
The framework and tools presented today provides me with a clear process for identifying and implementing needed changes based on data analysis.	48.4% (138)	44.2% (126)	7.0% (20)	0.4% (1)	0.0% (0)	0.0% (0)
The professional development received today is an improvement over professional development I have received in the past concerning data analysis.	57.0% (163)	32.5% (93)	9.4% (27)	0.7% (2)	0.0% (0)	0.3% (1)
The tools and process presented today have improved my ability to analyze and use data.	58.4% (167)	35.3% (101)	5.6% (16)	0.7% (2)	0.0% (0)	0.0% (0)
As a result of this training, I feel comfortable showing others how to access, analyze, and use data.	30.7% (87)	48.1% (136)	17.3% (49)	2.5% (7)	1.1% (3)	0.4% (1)
The tools and process presented today would enhance our school improvement planning process.	58.1% (165)	36.3% (103)	5.6% (16)	0.0% (0)	0.0% (0)	0.0% (0)
The tools and process presented today allow me to identify high achievement in other schools that we should investigate and possibly emulate.	62.7% (178)	33.1% (94)	3.9% (11)	0.0% (0)	0.0% (0)	0.4% (1)
The tools and process presented today will help me affirm a theory about student achievement I had before but could not quantify.	47.2% (134)	39.8% (113)	11.6% (33)	0.7% (2)	0.0% (0)	0.7% (2)
The tools and process presented today gave me a more comprehensive view of our program and instruction.	51.8% (147)	40.5% (115)	5.6% (16)	1.4% (4)	0.4% (1)	0.4% (1)
The process and tools presented today have convinced me that there is more to look at than just test scores.	61.2% (175)	35.0% (100)	3.8% (11)	0.0% (0)	0.0% (0)	0.0% (0)
The process and tools presented today will help me avoid unfounded assumptions when I analyze data.	51.6% (147)	40.4% (115)	7.4% (21)	0.7% (2)	0.0% (0)	0.0% (0)
The process and tools presented today will help me look at data while avoiding the bias of prior knowledge.	53.9% (153)	40.1% (114)	5.6% (16)	0.4% (1)	0.0% (0)	0.0% (0)
The process and tools presented today will help me find patterns in the data that are different from what I expected.	61.1% (174)	34.7% (99)	4.2% (12)	0.0% (0)	0.0% (0)	0.0% (0)

Appendix C
User Comments

SurveyResponses
Would you care to make any comments?
WOW!! This workshop needs to be brought to our school for one on one training! Can't wait to see individual student scores! Well done!!!
Wow! This is the best workshop I have ever been to. I can't wait to share this information with my superintendent, principal and teachers. There is so much information that can be used and it is so comprehensible. I love the way the information can be
Would love for all data to be this simple!
Would like to see year comparisons by strands that also includes sub-pops. This is a requirement for ACSIP. Great program. Much easier than Normes. Reminds me somewhat of the TinkerPlots Program that I use for data.
Would like to have a more in-depth training in using this model
Would like to be able to present the same data in different types of charts.
Wonderful program. Thanks for introducing this to us.
Wonderful and useful information, very user friendly. The charts and graphs and visual graphics make the data so much more understandable for teachers and administrators. Only had an hour and a half for this session. I wish I had a full day with traini
Way to go Neil!! Thanks, Sarah Alexander
very visual---easy to follow great deal of information available
Very user friendly!
Very user friendly and provides wonderful data.
very user friendly and informative to help guide instruction.
Very user friendly - easy to use and interpret
Very interesting program.
Very impressed with program. Once well-versed, this will prove an effective tool in data analysis.
Very good!! Provides lots of information in a user-friendly format. Look forward to using it in the future.
Very good program!
Unlike some other sites, this is very user-friendly. I will be looking for future training dates with hive/Neal Gibson.
Totally excited about the possibilities for data disaggregation that this website offers.
This was the greatest presentation of data that I have witnessed. I absolutely love the ease of moving throughout the program and the methods of data presentation. This will be a huge asset to our district. I can't wait to get back and share what I have
This will be a very helpful tool for me to use as we plan for professional development activities and intervention for our students.
This was the most "user-friendly", concrete representation of the data that I use to drive instruction, evaluate programs, and align curriculum.
This was a very helpful in-service with valuable information. I feel certain I will access this in the future. Thanks so very much for introducing me to the Hive. The class was very fast paced and I will definately go back and explore the many options
This system will allow more people to take advantage of the wealth of data at our disposal.
This system is awesome!
This seems easy to navigate and understand. I feel that teachers will be able to easily understand and navigate through this system.
This program will save me hours and hours of work as I compile the information legally required

SurveyResponses
Would you care to make any comments?
for our District Annual Report. Instead of using multiple sites, thumbing through mountains of paper work, or scrolling through pages of PDFs, this makes mult
This makes our school data less overwhelming and presents it in a manner everyone can understand and use. Great tool!
This is tremendously valuable tool. This is going to make the process of analysis more efficient.
This is the best user friendly system I have used for student data. Busy Administrators really need this.
This is the best program evaluation tool that I have ever used.
This is great!!! We need more intensive training and data needs to be given to Mr. Gibson sooner so the schools can get it into their hands sooner.
This is going to be very useful to me and to my district!
This is amazing! Thank you for putting this information together.
This is a user friendly way to look at data in a multitude of ways.
This is a great tool to show growth and to show weakness. As a teacher I feel I can benefit from this material.
This enables us to view an over-all picture that is easy to access for all. It is easy to understand and visual.
The instant charting is an excellent feature.
Thanks!
Thanks for all the hard work. This software takes data to another level and will allow for cleaner charts/graphs to use to improve student achievement.
Thank you Neal. This is going to make a much better tool to analysis data for individuals, teams, schools, & districts, especially for ACSIP planning. We can also use this to help choose students for Remediation, and staff hiring.
Thank you for working with the data available in order for teachers to readily utilize the data to create purposeful classroom activities, implement proven strategies and evaluate results.
Specialist needs the pass cods for all the districts that we are over for school improvement.
Pictures speak a million words about student achievement! I will be very anxious to get into hive and play with it.
Parents would not understand this data. I have trouble following some of the information. Could be structured in a "less analyzed" format.
Parent may have a difficult time interpreting and navigating but information is available for viewing.
Not at this time
Neal, Make a cheat sheet on what to chick to get what report.
LOVE the ease of access and co-op level information--would like to have co-op level information by strand.
Learned a lot, better than D2sc, navigation is better, better visualization
I'm not sure that I am familiar enough with the system to offer valid survey responses, but with my limited experience and understanding, I did try to complete each point.
I would like to see more detail in Strand analysis. The information that I saw in Quick Looks is not any different than what I am given in our ACTAAP reports. I believe that DETAILED Strand analysis is necessary to change instruction programs.
I would like to learn more about comparing the growth from second grade to third grade. Two different test and it is hard to see if the student growth drops or increases.
I would like to be able to compare our school district to other districts of similar demographics.

SurveyResponses
Would you care to make any comments?
Also, I would like to find some way to look at individual SLE performance of individual students on the benchmark results to see how a student performed on
I truly believe that this data tool can show adm/teacher/s that specific strategies/interventions applied in a class activities or with individual students can make a difference. Further I think with thoughtful consideration it can make a difference whe
I think this is a great tool to use.
I think this is a great system. It will be helpful in identifying students' needs, areas of concern for our building and finding other districts that could be of help to us.
I think the hive system could be useful to parents and to teachers who want to individualize, but we didn't get to look at that function....no problem...but I just don't feel competent to assess something I haven't seen.
I really like this!!!
I love this program. The brief presentation I saw opened my eyes about many things happening in my district.
I like the scatter plots that allow your to look at the different sub pops to see what the problems are. This was a very useful sight to me. I would like to be able to look at DIBELS scores off this site to look at sub pops to look at trends.
I have been enlightened. This is the most user friendly form of data presentation I believe I have ever had the opportunity to view. This WILL be used at our school to effectively drive our instruction and training.
I feel hive enables me to make sense of numbers that normally could be confusing in simple numeric form. Thanks for the work. It is really good.
I enjoyed the presentation. It will take a couple of staff development sessions to inform my principals, counselors, math coaches and literacy coaches on utilizing this information. Thank you very much, Tom Wilson
I don't remember the parent component.
I did like the simplicity of the program. It gives you an accurate picture of individual students which is very needed when making changes in the classroom.
I can't wait to have my curriculum teams trained in using this to diagnose problem areas for individual student intervention.
I came in the middle of training at director's request, so I'm just taking survey to see what it's about!
I am extremely impressed by the potential of this system; I just need time to become more familiar with it.
I am a strong visual learner and the visuals are strong. I can review these visuals with teachers and be able to generalize what areas that we need to work on.
Having the 2010 data in July will definitely be a benefit. We will be able to compare trend data and have current data to look at with ACSIP. The questions pose a good way to get lots of involvement from all staff rather than just presenting the data to
Great workshop. I can use all that I learned today. Thanks.
Great work on the data, Neal! This is the kind of site we need that, with minimum training, is VERY user friendly! Thanks for your work to help educators and students in Arkansas.
Great tool for school use to guide student instruction.
Great tool for analyzing data and making some sound instructional decisions!!!! Thanks!
Great source of data and easy to use.
great program and useful for teachers, adminsttrators and teachers for understanding growth, strands, and perceptions of how our school does compared to the state.

SurveyResponses
Would you care to make any comments?
Great Job! Love this software. Very user friendly!!!
Great job! Thanks for sharing with me. I really enjoyed the presentation.
Great I have already used it to help me with my reports
Great data resource! I am looking forward to digging in deeper and learning more.
good program
Exciting new program!! User friendly. Enjoyed the session.
Excellent training with huge data points to absorb and evaluate.
Excellent tool
Excellent software. Can't wait to share this data with teachers.
Excellent program that is time saving effective means of providing critical data to meet annual yearly progress. Please provide funding for Neal Gibson to continue his efforts to meet the needs to educators.
excellent program and useful!
Excellant data, very descriptive, and easy to use.
Easy to use and lots of information that is visual
Eager to be able to use the information presented in today's presentation.
Can't wait to see more of this!
Because of the short time allocation and problems within the lab our time was limited. At this time, I don't know enough about this program to make an educated comparison. I am very interested in this program and will work with it at a later time. Whateve
Awesome!
As a high school principal, I will use this......
Am anxious to learn more and practice with the program.