Teacher-Generated Final Exams in High School Science: Content, Rigor, and Assessment Literacy

By

Michael Lach
B. A. Carleton College, 1990
M. A. Columbia University, 1995
M. A. Northeastern University, 2000

Thesis

Submitted as partial fulfillment of the requirements for the degree of Doctor of Education in Educational Leadership in the in the Graduate College of the University of Illinois At Chicago, 2014

Chicago, IL

Defense Committee

Steve Tozer, Chair and Advisor
Jim Pellegrino, Learning Sciences
Samuel Whalen
Chris Miller
Don Wink, Chemistry

Acknowledgements

I am very grateful for the efforts of my thesis committee, and in particular to my advisor Steve Tozer. You all demonstrated an incredible amount of patience, as this has been nearly a ten-year effort as I've taken on various roles in various cities during that time. And of course, your guidance and insights were an essential piece that enabled me to pull this together.

I am also grateful to my CPS supervisors over this timespan--Xavier Botana, Arne Duncan, Barbara Eason-Watkins, David Gilligan and especially Marty Gartzman—for letting me pursue this research while simultaneously helping the district's efforts.

Thank you also to my family, particularly my wife Leslie Darling and my dad Joseph Lach.

The team at Loyola University's Center for Science and Math Education helped with the *Survey Of Enacted Curriculum* coding work, which formed the core of this analysis.

This work is for and about the teachers of the Chicago Public Schools, and the past, present, and future members of the Office of Mathematics and Science. You've taught me so much, and it continues to be an honor to work and learn alongside you.

Table Of Contents

List Of Tables

List Of Figures

Summary

This study investigates a large collection of teacher-generated end-of-semester final exams from Chicago Public School high school science classrooms in order to explore the depth and breadth of content that students learn in science classrooms. Teachers focus on a specific set of scientific content that is driven by district guidelines and popular textbooks but not particularly aligned to standards. To most teachers, rigor means coverage instead of intellectual press. The assessments, while unsophisticated, seem to be delivering what is expected of them—a way to mimic the most basic format of the ACT exam quickly. There was little variation among high poverty and low poverty schools, matching national data and indicating issues that are more due to a particular culture of science teaching and learning than driven by particular contexts. The study identifies implications for the observed homogeneity of final exam rigor and content, identifies gaps between how the routine of final exams are design and implemented in schools, and discusses similar methodological efforts that could enhance the ability of schools and districts to access useful information about the technical core of instruction.

Chapter 1: The Problem

The Problem

In science instruction, breadth and depth of learning both matter. New scientific discoveries are made on a daily basis, so the amount of scientific knowledge is continually increasing. K-12 teachers and school leaders know all this new knowledge must be moderated in the classroom—an introductory biology course cannot focus solely on the latest genetic discoveries, for instance. At the same time, students need to learn both the unique ways of knowing associated with the scientific endeavor as well as sufficient content to transcend the often-naïve misconceptions that are developed in their everyday interactions with the natural world. This means studying particular topics with depth.

The rise of standards-based reform over the last few decades in American education created a wave of cascading policies and changes—and is poised to do so again with the recent advent of the *Next Generation Science Standards* (NGSS Lead States, 2013b). Yet the heart of this work—the teaching and learning—remains “loosely coupled” (Weick, 1976) within classrooms, despite all the attention and energy that have focused upon it. Richard Elmore explains the “bane of ‘loose-coupling’” in *Building A New Structure For School Leadership*.

> *"…The 'technical core' of education—detailed decisions about what should be taught at any given time, how it should be taught, what students should be expected to learn at any given time, how they should be grouped within classrooms for purposes of instruction, what they should be required to do to demonstrate their knowledge, and, perhaps most importantly, how their learning should be evaluated—resides in individual classrooms, not in the organizations that surround them" (Elmore, 2000, page 6).*

Tools that enable us as both researchers and practitioners to explore this technical core at scale have the potential to be quite powerful. To investigate the depth and breadth of content that students learn in science classrooms, one relevant artifact is the summative exams that teachers write and students complete at the end of each semester. These "final exams" are hallmarks of the American high school, and ostensibly address both the depth and breadth of content contained within a course. And while individual teacher-written exams likely vary greatly on all sorts of dimensions, analysis of large collections of such exams can reveal patters than describe the depth and breadth of science learning expected of Chicago high school students.

In the 2003-04 and 2004-05 school years, several hundred final exams from high school science classrooms were collected as part of a school improvement strategy initiated by the Chicago Public Schools Chicago Math & Science Initiative. In studying a collection of these artifacts of individual classroom practice, I have explored an important part of the technical core of high school science teaching and learning from a unique vantage point. This data creates a picture of a unique routine within high school science teaching, and has the potential to help explain what high school science teachers in Chicago value about the content they teach, how they organize that content, and the sort of demands they expect of their students. This study describes and analyzes that data and paints the resulting picture of practice.

Why Is This A Significant Problem?

There are several reasons this is a problem worthy of study. Most importantly, the technical-core issue described above is one that is intractable from the practice of educational leadership and attempts to improve student learning at scale. Particularly from the district and state level, obtaining meaningful access to the thinking, decision-making, and practice of classroom teachers remains difficult to do. Solutions that address this challenge will certainly help aid understanding and bring clarity to the work of district and state level policy makers.

At least six other reasons describe why this is a significant problem to tackle.

First, the semester and final exam in high school is part of the culture of American education, particularly in science. Teaching is a cultural activity, and the decisions teachers make depend on the school and educational culture in which they operate (Stigler & Hiebert, 1999). High school science has a few touchstones in our society—the frog dissection, the mad scientist chemistry teacher who “blows things up,” and onerous lab reports, to name a few examples—and a summative final exam is one of them. While some (Alfie Kohn, Deborah Meier) have valid concerns about the purpose and usefulness of end-of-course assessments, in practice, they have been institutionalized for decades. For this reason alone, studying them is important.

A second reason is the particular way of knowing in science: the topics that are taught matter, but the field of science is one defined by knowledge of both process and content. The standards movement has expended great energy to articulate the body of knowledge and scientific ways of knowing that are relevant for students and teachers at the K-12 level (Bybee, 1997). Important work strives to continually update them (Quinn, Schweingruber, Keller, & Krone, 2010), and the latest *A Framework For K-12 Science Education* talks about "3-D science" where practices, cross-cutting concepts, and disciplinary core ideas are intertwined (National Research Council, 2012). The American mathematics and science curriculum has been derided as "a mile wide and an inch deep" (W. H. Schmidt, 2005). However, to think that just because particular documents describe topics for study means that they are actually taught is a different matter entirely (for instance, Spillane, 2004). Additional insight about what teachers really teach and how decisions about particular content are made will help school and district leaders, assessment designers, and curriculum developers better understand what actually happens between teachers and students in science classrooms.

A third reason is the more recent groundswell of dialogue about rigor in high school coursework. Many influential national groups—the Bill & Melinda Gates Foundation, the American Diploma Project, the American Youth Policy Forum, the College Board, to name a few—have recently produced reports that decry the current lack of rigor in American high school classrooms. The Obama Administration has made "higher standards" a focus of its school reform agenda (Duncan, 2010), and "college and career readiness" is now part of the educational reform lexicon. (See http://www.achieve.org/college-and-career-readiness.)

These policy proposals and initiatives have been echoed by similar findings from researchers. A study released by the U. S. Department of Education, for example, found that "the academic intensity and quality" of a student's course of study was a far more powerful predictor of bachelor's degree attainment than class rank, grade point average, or test scores (Adelman, 1999). The Chicago Consortium on Chicago School Research's work to develop the idea of "on track" indicators connects strongly to conceptions about rigor and academic press for students (E. Allensworth & Easton, 2005).

A fourth reason to explore questions about content and rigor in high school science classrooms is that some science content is controversial. Debates about evolution and creationism regularly make front-page headlines. Controversies about environmental science issues (the National Science Teachers Association's decision not to distribute free copies of Al Gore's *An Inconvenient Truth* DVD for fear of upsetting some constituencies, for instance) or evolution (Bill Nye's recent debate at the Creation Museum was front-page news) are essentially debates about what content students should students learn. And while there is general consensus among science educators that topics such as inquiry and the nature of science are important (Dow, 2000), there is compelling evidence that teachers do not understand how this sort of content can or should be included in their coursework (Banilower, Boyd, Pasley, & Weiss, 2006). To date, there is not a clear picture of how science teachers select the topics that they are to teach, and how they use the context of district and state standards to do so.

A fifth reason to explore these questions is the phenomena articulated by researchers such as Tyack and Cuban (Tyack & Cuban, 1995) and Spillane (Spillane, 2004) that describe how reforms get assimilated into actual classroom practice. Messages from central office or the local administration are often misunderstood or misapplied in the classroom setting. An understanding about what happens in the actual classroom, and the decisions that teachers make in response to a new idea or new reform, will help leaders and policy-makers in their efforts to transform schools. At the system level, district-wide assessment data can tell the outputs and results of the educational system, but identifying causes for changes in performance and the possible interventions that will enable greater effective-ness remains difficult.

A sixth reason to explore these questions is that assessment literacy matters. "Data-driven decision making" (Shirley & Hargreaves, 2006) is now also part of the American ed-ucational lexicon. It is clear that in many districts, teachers and schools have struggled to make sense of the new student assessment data that is available (Goertz, Oláh, & Rigan, 2009). By exploring the thinking and decisions teachers use when creating their in-class assessments, a picture of what teachers know and do not know about assessments can be established, providing the foundation for further work and study.

Conclusion

In this chapter, I have described the problem of understanding at scale the decisions high school science teachers make about the content and rigor they embrace in their clas-ses, and explained why it is a worthwhile problem to consider. In chapter 2, I will review the literature base to determine what we know and do not know about this topic.

Chapter 2: Literature Review

Introduction

In the previous chapter, I discussed the importance of three interrelated issues about science teaching in Chicago high schools: what do high school science teachers in Chicago value about the content they teach, how do they organize that content, and what sort of demands do they expect of their students? There are several different collections of scholarship about high school science teaching and learning in the United States in general and in particular in Chicago. In this chapter, I will provide an overview of those collections of research literature. In Chapter 3, I will describe some of the gaps in this literature base, and how this study helps to close that gap.

Organizing The Literature Base

I will begin with an analysis of what we know about high school science teaching, paying particular attention to how teachers react to policy mechanisms such as standards and assessments. There is a gap here between what practices research tells us work, and the outcomes we observe, however, meaning that our understanding of practices at scale is limited.

I will next ground the work in Chicago and describe the particular context into which my research is situated. I will draw from district data, the work of the Consortium on Chicago School Research at the University of Chicago, and the extensive research and evaluation background of the Chicago Math & Science Initiative (CMSI). The latter includes research papers produced internally by the district, and the University of Illinois at Chicago (now Loyola) PRAIRIE group, often as part of National Science Foundation funded efforts. While these papers set some context and define in part how individual teachers and schools interact with their science content and their students, there is little here that describes the routine work of science teachers.

The next body of literature focuses on science content—both breadth and depth—as typically defined for high school students and teachers. This includes various sets of national and state standards, the design work that forms the basis for curriculum development, and work about "learning progressions" and student misconceptions within the various scientific disciplines. This literature will address various definitions of academic press and rigor, but most of it describes the state to which teachers, schools, and curriculum writers should aspire, not the level of current classroom practice. I will pay special attention to some particular content areas that have received lots of attention for various reasons over the past decade.

The fourth body of literature describes assessment use by teachers, the connections between large-scale and classroom level assessments, and the decision-making and policy implications that entails. This body tends to be rather theoretical—amid the calls for more "data driven decision making," there are few examples that show how this can be done at scale in a fashion that lives up to the rhetoric.

Connecting all of these bodies of literature is a final group that discusses the leadership and policy issues at play. This is important because the practice and craft of teaching does not occur in a vacuum, but in schools described by routines and overseen by leaders, governed by policies.

High School Science Teaching

There is considerable research that attempts to describe effective science teaching. Numerous policy efforts have attempted to create science "master teachers" (White House Office Of The Press Secretary, 2012) and define the knowledge, skills, and dispositions needed for exceptional practice (The National Board For Professional Teaching Standards, 2003). The *Handbook Of Research On Science Education* (Abell & Lederman, 2007) synthesizes much of what we know about teaching, learning, and leading science at the K-12 level, though as part of the inevitable process of synthesizing many studies the local contexts are downplayed.

Some more recent and targeted studies attempt to highlight both practices that are particularly important and the processes by which teachers learn or adopt these practices. For instance, a recent research synthesis shows that research "indicate[s] a clear, positive trend favoring inquiry-based instructional practices, particularly instruction that emphasizes student active thinking and drawing conclusions from data" (Minner, Levy, & Century, 2010), thought it is clear that these practices are difficult to enact successfully in many classrooms (Tang, Coffey, Elby, & Levin, 2009). Some studies have explored how science teachers respond to standards and assessments and accountability provisions, but these tend be based on either small numbers of teachers in focus groups (Donnelly & Sadler, 2009) or tied to specific interventions, such as instructional materials (Ball & Cohen, 1996) or concept substitution (Grayson, 1996).

High School Science Teaching In Chicago

These questions highlighted at the beginning of this chapter are grounded in the political, organizational, and educational context of Chicago. As Kahle comments when reviewing large-scale attempts at science education reform, “success depends primarily on local factors, not federal guidelines or funding levels” (Kahle, 2007). Much has been written about high school performance in the Chicago Public Schools. In addition to the variety of data from the district (Bugler, 2006), the Consortium’s regular surveys and reports (Luppescu & Hart, 2005) provide a way to benchmark and ground progress.

The Chicago Consortium On School Research

Essentially, overall student achievement in Chicago high schools has increased slightly in the past decade (Luppescu et al., 2011). There is a gap between college aspirations and actual attainment, as students do not leave high school with the knowledge, skills, and dispositions to be successful in higher education (Roderick et al., 2006). A leadership focus seems to be on preparation for the Prairie State Achievement Exam (PSAE), the high-stakes exam for high schools (Stoelinga, Hart, & Schalliol, 2008), though this “test-prep” focus does not provide very strong student learning results (E. Allensworth, Correa, & Ponisciak, 2008b). Small high schools have greater amounts of program coherence than large high schools, but student outcomes are generally similar to large comprehensive high schools (Kahne, Sporte, de La Torre, & Easton, 2006). When professional communities exist in small high schools—not a given—they tend to focus on supportive practices rather than developmental ones, and the daily demands of teaching often made teacher collaboration difficult (Stevens & Kahne, 2006). On the PSAE, performance is low, particularly in science (Bugler, 2006), though scores on one subject are associated with low scores on the others (Ponisciak, 2005). Graduation rates in Chicago are low but increasing slightly, but the variance is high between schools (E. Allensworth, 2005). In the early redesign work sponsored by the Gates Foundation, despite increased optimism and personalization, implementation and operations hurdles eclipsed instructional issues as the primary focus of schools (Sporte, Kahne, & Correa, 2004), though later efforts began to show some signs of academic progress (Sporte, Correa, Hart, & Wechsler, 2009). Together, the portrait these paint of Chicago's high schools is one with lots of energy and enthusiasm for reform, but to date, not enough progress.

Mathematics and Science In Chicago

There is a host of literature about mathematics and science performance in Chicago as well, thanks to the program evaluation efforts of the Chicago Math & Science Initiative (CMSI). The evaluation and research plans are presented in annual documents describing the strategies and approaches used to collect, analyze, and disseminate data in 2003-04 (Feranchak, 2003), 2004-05 (Feranchak & Price, 2004), and 2005-06 (Feranchak, Price, & Lemke, 2005), and most of the outcome data mirrors the broader reports referred to above. (The progress at the K-8 level has been more pronounced.) The annual reports to the National Science Foundation describe strategies and results annually (Lach, 2006). It is clear that capacity for content-based change remains limited, as efforts to increase the leadership capacity of central office staff members in mathematics and science have had mixed results (Hallman, Fendt, & Wenzel, 2003). As time has passed, those efforts did appear to increase the district-wide instructional coherence, but capacity remained thin (Hallman, 2004). A review of teacher qualifications shows most high school teachers have the appropriate content background for the courses they teach, with key gaps in the physical sciences and special education (LaForce, 2004). It is also clear there are few commonalities among instructional materials used in CPS high school science classrooms (Deiger, 2003), and the district lacks a coherent policy or plan around course enrollment and sequencing (LaForce & Feranchak, 2003). At the high school level, there is an emerging sense that the college-prep focus likely has not made much impact for many students (Mazzeo, 2010). Changes in high school graduation requirements were made in the early 1990s, dramatically increasing science course taking requirements. These changes did not help students learn more science and actually may have hurt their college prospects (Montgomery, Allensworth, & Correa, 2010).

Much of this literature—particularly the Consortium reports—describes and benchmarks the district's performance. When policy levers are considered, they are applied at a broad scale, and there is little insight into how the various policy levers connect to one another at the classroom level. The CMSI work also mostly provides benchmarking, though it does delve into the workings of the central Office of Mathematics' and Science. As Spillane has described in *Standards Deviation,* the connections between different levels of the educational system often dilutes the implementation of instructional innovations (Spillane, 2004), so probing how the various levels connect and impact the practice of teachers will be important. The literature discussed above generally describes the context and the setting, but does not really explain what this looks like in teacher classrooms around the district, or how various policies influenced classroom instruction and student learning.

Science Content, Science Rigor

Now, I will describe several bodies of literature that address nature of science content and rigor for high school classrooms. This includes an explanation about how decisions are made concerning science content, an overview of the current status of state and national standards for science, the nature of learning progressions, and various definitions of rigor. I will also tackle some important "special cases" of science content that have particular relevance: environmental education, including climate change; evolution; and inquiry.

Drivers Of Science Content In Classrooms

It is clear that the primary driver of science content in classrooms is the textbook that teachers use. Yager and Stodghill commented back in 1979 that:

> *"Science in the school program can be characterized by one word—textbooks. The science curriculum exists as the facts and concepts that are traditionally packaged in textbooks. The textbook not only determines the content, but the order, and the examples, and the application of that content." (Yager & Stodghill, 1979)*

This is still the case, as reported by Lyons as part of the 2012 National Survey of Science and Mathematics Education (Lyons, 2013).

At the national level, there is substantial consensus regarding what science high school students should know and be able to do. Two sets of national standards have existed for some time, the *National Science Education Standards* (National Research Council, 1996) and the *Benchmarks for Science Literacy* (American Association for the Advancement of Science, 1994), and both have paved the way for subsequent systemic reform and policy decisions (Kahle, 2007). While more similar than different, the *Benchmarks* include a broader vision for science literacy than the National Academy's version (Bybee, 1997); either sufficiently defines what science students should know and be able to do, and both were influential in crafting state science standards. Building on these documents, the National Academies recently developed a *Framework* for science education (National Research Council, 2012) which was used by Achieve, Inc. and some state partners to create the *Next Generation Science Standards* (NGSS Lead States, 2013b) which as of early March 2014 have been adopted by nine states. Illinois is poised to adopt them as well (Heiton, 2014).

Evidence is emerging that by focusing more on depth than breadth, as the NGSS are designed to do, is important in high school for performance in science in college. As Schwartz, *et. al.* report in their study relating the performance of college students with high school experiences,

> *"Students who reported covering at least one major topic in depth, for a month or longer, in high school were found to earn higher grades in college science than did students who reported no coverage in depth. Students reporting breadth in their high school course, covering all major topics, did not appear to have any advantage in chemistry or physics and a significant disadvantage in biology" (Schwartz, Sadler, Sonnert, & Tai, 2008).*

While we have good data that shows Chicago's high school science performance is not that strong, we do not have a very robust understanding of how ideas about content and rigor get operationalized at the classroom level by teachers.

State Standards

Information about science content and performance in Illinois is limited. Information about high stakes testing in Illinois, including the technical manual for the Prairie State Achievement Exam (PSAE) (Illinois State Board Of Education & ACT, 2007) and the accompanying science overview for teachers (Stanko, 2007), is readily available, but focus mostly on the psychometrics of the test design and logistics for implementation. During the 2003-04 and 2004-05 school years, the high school PSAE consisted of two components spread over the two days of testing: the ACT science reasoning exam on day one and an ISBE-develop science exam on day two (Lach, 2004). According to the ISBE Science Assessment Framework PSAE Grade 11, day one focused 85% on scientific inquiry and day two focused 75% on the content standards (goal 12) (Illinois State Board Of Education, 2006). And while both statewide and Chicago PSAE results are public (Bugler, 2006), and some reviews of PSAE results exist (for instance, Ponisciak, 2005), there are few deep efforts to probe at what is really going on in classrooms with any grade or content specificity. (There are no statewide indicators of progress in specific science courses or grade levels, like Biology or 9th grade science, for instance.) Relatively little information exists about the statewide contexts for science, technology, engineering, and mathematics (STEM) education reform, though a remarkably noncommittal overview is available from the Illinois

Business Roundtable and Northern Illinois University (Northern Illinois University, 2006) and recent attempts have been made to set an agenda for teachers (Jackson *et al.*, 2013) and the state (C-STEMEC, 2013). Some data does exist about college readiness and performance, showing that only one-third of the high school class of 2002 graduated ready for college (Presley & Gong, 2006), though there is little content-specific data there, either. Statewide, the teaching force quality is shown to be concentrated in areas of relative wealth, and that the influence of high quality teachers is strongest at the high school level (Presley, White, & Gong, 2005). Specifics about teacher preparation in particular content areas such as mathematics and science are missing.

Certainly, the Illinois State Standards for science (Illinois State Board Of Education, 1985) and the accompanying performance descriptors (Illinois State Board Of Education, 2001) will be cornerstone of any effort to gauge the rigor and academic press of high school science classes—and as mentioned previously, new standards will likely be adopted in 2014. A rather cursory review of current standards is available, and places Illinois in the middle of the pack when compared to other states (Gross *et al.*, 2005). In Illinois, there is no data about performance on particular courses in high school, and the connections between statewide and district policy and the classroom are loose at best.

Learning Progressions

The emerging discussions about learning progressions (Committee on Science Learning Kindergarten through Eighth Grade, Duschl, Schweingruber, & Shouse, 2007) describe how an understanding of conceptual flow within a particular aspect of a discipline can help teachers identify the important ideas when teaching (BSCS, 2006). Studies have connected this work with curriculum design (for example, Roseman, Caldwell, Gogos, & Kurth, 2006), but only minimal work has shown how it can emerge in classroom level assessments via specific instructional materials (for example, Stern & Alhlgren, 2002). Understanding how science teachers decide which content to assess has the potential to inform these discussions.

Rigor

The policy airwaves have been filled lately with cries for more academic rigor. These reports tend to be a mix of national test score analysis and political commentary, and often compare high school work with that accepted at colleges. A report by ACT, Inc. shows that course taking is no longer the indicator it once was for college success, as the core content of courses is often ambiguous or watered down (ACT Inc., 2007). Many of the course titles of advanced courses do not reflect advanced topics (Dougherty, Mellor, & Jian, 2006). Furthermore, many high school courses do not align with content and rigor expectations of colleges (Venezia, Krist, & Antonio, 2003). As mentioned above, there is evidence suggesting that depth of content leads to better postsecondary outcomes (Schwartz et al., 2008). While advanced placement course do provide some level of specificity about what expectations make sense for students, and some initial work has done to broadly define what rigor looks like in effective high school courses (ACT Inc. & The Education Trust, 2004), this literature provides more impetus for the proposed research than it focuses the picture of classroom practice. ,The College Board has since launched an effort to redesign their science course offerings (National Science Foundation, 2006).

Special Topic: Environmental Science and Climate Change

Given City Hall's interest in environmental issues, and the national attention placed on climate change, the focus on environmental science in Chicago has never been stronger (Ferkenhoff, 2006). This comes amid somewhat of a complicated time for environmental science at the national level, as the political debates sharpen. For instance, the National Science Teachers Association recently found itself in hot water for refusing to freely distribute copies of Al Gore's award winning *An Inconvenient Truth* for fear of upsetting corporate sponsors (David, 2006). As a discipline, the understanding is now emerging that environmental education has its own unique epistemology due to its connections to the social aspects of science (Hart, 2007). A description of environmental science as an interdisciplinary, contemporary, and applied science at the high school has been articulated in both the national literature and as one component of the Chicago Public School's high school improvement efforts (Edelson, 2007), but it is unclear what is the most common type of environmental science practiced within the Chicago Public Schools.

Special Topic: Evolution

No science education issue has attracted more attention in recent memory than the study of evolution. High-profile court cases ("Tammy Kitzmiller versus Dover Area School

District," 2005) have fueled the controversy, and some studies indicate many high school biology teachers merely ignore it as a topic all together (Monastersky, 2005). No available data describes the degree to which evolution is taught in the Chicago Public Schools.

Special Topic: Inquiry and the Nature of Science

Inquiry is another science topic that has been the subject of much misunderstanding. While the national standards delineate inquiry as separate from the nature of science (Dow, 2000), the degree to which science teachers make the distinction remains unclear (N. Lederman, 1998). In the *Next Generation Science Standards*, this idea is included in the “practices” section and integrated into all the standards (NGSS Lead States, 2013b), though the designers and standards authors note the challenges this will mean for implementation (Board On Science Education, 2014, in press). The current Illinois state standards do not make the distinction very clearly either, as inquiry and technological design are grouped in one goal (goal 11), with what could be most closely aligned with nature of science in another (goal 13) (Illinois State Board Of Education, 1985). Inquiry is not well tested on the state’s high-stakes exam (Illinois State Board Of Education, 2005), so reviewing what happens in classroom level assessments about inquiry will fill a sizable gap in our understanding of science instructional practice. Evidence suggests that teachers with stronger content knowledge background, and more professional development that is firmly grounded in content, use inquiry-oriented teaching approaches more frequently than other teachers (Smith et al., 2007), though this study relied on teacher self reports about methods and techniques to measure impact on instruction.

Assessments and Assessment Systems

The literature base on science assessment is extensive, beginning with the National Academy's *Knowing What Students Know* (Pellegrino, Chudowsky, & Glaser, 2001). In this important work, the authors describe how cognition, observation, and interpretation form the foundational elements of all assessments, and describe how careful consideration of these three components of an "assessment triangle" would result in assessment systems and practices dramatically different (and more effective) than currently in mainstream use. *Systems for State Science Assessment* (M. R. Wilson & Bertenthal, 2005) attempts to take these ideas and describe what a large-scale science assessment system should look like, in advance of the No Child Left Behind science testing mandate for the 2007-08 academic year. The classroom assessment supplement to the National Science Education Standards (Atkin, Black, & Coffey, 2001) describes the connections between research and teacher practice around classroom-based assessment, and documents clearly how teachers can incorporate stronger assessment strategies into their classrooms. Another report from the National Academy of Sciences, *Assessment in Support of Instruction and Learning: Bridging the Gap Between Large-Scale and Classroom Assessment* (Committee on Assessment In Support of Instruction and Learning, 2003), makes concrete the connections between the classroom level and large scale exams. It links to much of the synthesis work in other National Academy reports in order to discuss how classroom assessments and the management of them fit within an overall system of assessments, and highlighting the curricular and instructional decisions that can be made to improve local level teaching and learning. Recently, the National Academies have synthesized these recommendations in light of the recently released *Next Generation Science Standards* (Committee on Developing Assessments of Science Proficiency in K-12, 2013).

Another body of literature has emerged addressing comprehensive assessment plans for districts. Much of this is in response to the assessment and accountability provisions in the No Child Left Behind legislation, calling district-level assessments ultimately “a relatively weak intervention, because while it reveals shortcomings, it does not contain the guidance and expertise to inform response” (Supovitz, 2009). Additional work has explored using formative assessment mechanisms as a way to increase overall capacity. For instance, the National Center for the Improvement of Educational Assessment has released a framework (Perie, Marion, & Gong, 2007) to help state and district leaders make decisions about interim and formative assessments—mid way between the classroom and large-scale assessment systems, and in sync with the National Academy focus on a comprehensive and aligned assessment system.

Since the National Academy reports both synthesize the existing literature and include commissioned white papers to fill in identified research gaps, they provide a solid launching pad for additional research and study. The literature cited above generally describes high-level policy work and classroom level specifications—the intended audiences of both the *Inquiry And The National Science Standards* (National Research Council, 2000) and the *Classroom Assessment And The National Science Standards* (Atkin et al., 2001) supplements are teachers. The studies that are cited include descriptions of hothouse best-practice designs from much of the instructional materials development work, like the *SEPUP* assessment system (M. R. Wilson & Draney, 1999), but little data about what actually is happening in classrooms around assessment in science, particularly at scale. To create a complete picture of the current state of assessment at all levels, more insight into teacher practice is needed.

There is an emerging body of evaluation literature that describes how attempts to create large scale formative or benchmark assessment systems have fared. In Philadelphia, implementation of an aggressive benchmark assessment system was limited by local school resources and capacity (Bulkley, Christman, Goertz, & Lawrence, 2010). In Providence, a "well-crafted system of quarterly assessments" was implemented and later discarded, as the benefits never quite materialized from the design (Clune & White, 2008). Even the Consortium has moved to claiming interim assessments as "easier said than done" (Goren, 2010).

<u>Scale, Implementation, and Leadership</u>

There is only limited data on large-scale improvement in mathematics and science. And Cohen and Ball comment, "like love affairs, many [large scale] innovations consist more of hopes and dreams than the carefully designed details of daily operations that often are required to make appreciable change" (Cohen & Ball, 2007). The ARC Center's tri-state study (Sconiers, Isaacs, Higgins, McBride, & Kelso, 2003) shows that curriculum-driven instructional improvements can happen at a large scale—the changes in teacher practice brought on by well supported curriculum materials result in new practice and student learning gains. This is matched by several meta-studies that describe even greater impact from curriculum-focused intervention efforts (Whitehurst, 2009). After five years of implementation, the National Science Foundation local systemic change grants had only mixed results (Kahle, 2007), but the ten-year capstone report shows that both pedagogical change and student achievement improvements do occur via curriculum-centric reforms (Banilower et al., 2006), though teacher turnover and administrative support remain the primary challenges (Banilower et al., 2006). Several international studies, including those based on the Programme for International Student Assessment (PISA) (Watanabe & McGaw, 2003a) and the Third International Mathematics and Science Study (TIMSS) (Gonzales et al., 2004) data, include powerful insights into the role of curriculum (Watanabe & McGaw, 2003b) and support structures in advancing student achievement as well, particularly as related to opportunities to learn (W. H. Schmidt, Zoido, & Cogan, 2014). Essentially, the message of these reports is that curriculum matters, and matters in ways that are unique and particular to mathematics and science improvement.

In *Transforming Teaching In Math and Science* (Gamoran et al., 2003), the authors describe some of the district and school level structures that impact mathematics and science learning. In presentations made based on this material (Gamoran, 2004), it is clear that a focus needs to be on “discovery teaching” (Hammer, 1997) or “teaching for understanding,” which by definition do not work very well in routines or procedures grounded in restrictive organizational cultures, and probably is not included on most state standards (Finn, Petrilli, & Julian, 2006). Gamoran argues that schools need to expand the typical focus on “a predictable flow of material resources” to a broader focus on material, human, and social resources. The inherent uncertainty in discovery teaching speaks to a need for more organic and less bureaucratic management, and leadership based on “expertise and commitment, not just position.”

What does that mean for the practicing high school science teacher? Gamoran's work takes a science and mathematics focus to some of the work on professional learning communities (Dufour & Eaker, 1998) and generally progressive-minded school reform ideas, with a healthy dose of 2000-style business management thrown in. Both Dufour and Tim Kanold, who followed DuFour as superintendent at Stevenson High School and then later as a professional learning community advocate, used semester and final exams as the "organizational routine" (Spillane, 2006) to initiate and sustain their professional learning communities (Kanold, 2003). Halverson describes how leadership artifacts "such as policies, programs and procedures" are the heart of leadership practice and can be used to create the sorts of schools that Gamoran envisions, and that the TIMSS and PISA assessments demand (R. R. Halverson, 2003). So exploring the routine of semester and final exams will provide insights into both teacher practice and leadership at the local school level.

The systemic change literature will also certainly be important. While much of Elmore's thinking has been captured in both the writing of Gamoran and Spillane, he remains a giant in the field. In *Bridging the Gap Between Standards and Assessment* (Elmore, 2002), he argues for the importance of professional development and general transparency, as schools in the United States tend to isolate and ignore the "technical core" via structures of loose-coupling. I suspect that as the content and curriculum in mathematics and science classrooms becomes more transparent via an analysis of teacher's values and assessment practices, issues of school leadership and organization will follow. Connecting these will have major systemic implications, perhaps involving the way schools are organized or the preparation of teachers and principals. This literature will be helpful in identifying those sorts of changes.

Chapter 3: Conceptual Framework

Chapter Introduction

This chapter describes the conceptual framework that will be used to organize this research project. Beginning where I left off with the previous chapter, I will identify what the literature base does not tell us. From there, I develop an approach, which leads to the research questions. Following those questions, I discuss the data set I will use to generate findings, as well as how my own background influences the issues at play here. In Chapter 4, I will describe the methodology used to collect and analyze the data for this research project.

What the Research Literature Does Not Tell Us

In chapter 2, I provided an overview of the existing literature about Chicago education reform, with a focus on mathematics and science improvement efforts. I provided some national perspectives on high school science teaching and learning, described the current knowledge base about science assessments and assessment systems, and the history of national science standards. I reviewed the literature on science standards implementation, rigor, and impacts at scale.

Two related issues emerge from this literature review. First, it is clear that there is a gap in the literature base when it comes to describing both the content and the level of rigor in Chicago high school science classrooms, and the processes by which schools and teachers make decisions about which content to cover, the level of rigor to demand, and the nature of the tasks to assign. In this research, I will explore a series of questions to advance our understanding of these assessment dimensions. Second, this gap and this context both imply a particularly useful theoretical perspective that enable a unique approach to answering these questions.

I will address each of these in turn in the next sections.

Gaps In The Research Literature

The research base provides clear guidance on the macro-level outcomes of high school science teaching, learning, and leading within the Chicago Public Schools. The message is clear: student scores are low, performance is not good enough. But other than broad-brush statements, there is little understanding why performance remains where it does. Policy makers and leaders talk about poor leadership, uneven teacher quality, incoherent systems, and poor funding, but their explanations rarely dig below such surface statements to provide a more comprehensive and nuanced explanation for why the performance is so low. At the school and classroom level, the various academic and content standards provide a clear sense of what performance should look like, but not what practice currently looks like or how decisions about content are made—and there is clear evidence from the literature that teachers do not implement standards they way most policymakers intend them to (Spillane, 2004). The literature does not describe what science teaching practice looks like within Chicago high school science classrooms, other than the results as captured in specific and rather limited standardized assessments. It does not create a picture of practice sufficient to connect this performance to teacher decisions, school and leadership supports, and district-wide performance outcomes. Answering those questions—about the content teachers focus on in their classrooms, and the level of intellectual demand they use in their classrooms, and teachers' rationales for decisions about both—will be the focus of this research.

An Approach

The gap in the literature base described above also highlights a promising approach to investigating these questions, providing a warrant to address these questions from a particular theoretical perspective. My working premise is that teacher-generated end-of-semester high school science exams constitute important evidence of the content and intellectual demand that teachers value in their classrooms. Collected broadly, these exams will provide a great deal of information about the content, rigor, and assessment practices of CPS science teachers across the district.

What are high school end of semester exams? In the Chicago Public Schools, they are school required, teacher generated, summative assessments designed to assess learning of course content that are given twice each course, at the end of each semester. Exams such as these have a long history and tradition in American education. An example of an exam from more than a century ago was recently released and was "apparently a big deal" for all those involved (Bullitt County Geneological Society, 1912). It is reasonable to believe that virtually every educator active today experienced final exams as a high-stakes assessment in high school and in college; they were ubiquitous throughout the 20th century and into the 21st. In Chicago, while the precise degree of accountability generally varies by classroom, high school exams are clearly still "a big deal"—counting for a sizable portion of a student's grade for the course.

Why can looking at teacher-generated end-of-semester high school exams tell us anything meaningful about instruction, schools, and schooling? There are several reasons.

It Is Part Of CPS Practice

Organizational routines involve “a repetitive, recognizable pattern of interdependent actions, involving multiple actors” (Feldman & Pentland, 2003). My years of experience with the district shows that end-of-semester grades and final exams are an important part of the district routines—to schools they are a “big deal.” Schedules are changed, meetings are held, and energy is expended at all levels of the school in the hustle to write, administer, score, grade, and issue grades and end-of-semester report cards. This makes sense—at the high school level, the goal for students is to accumulate graduation-necessary course credits, which are issued each semester by teachers (Chicago Public Schools, 2004). As a teacher, I was expected to write and administer an end-of-semester final exam in all my courses. As an administrator, many of our improvement strategies and professional development agendas focused on assessment and the end-of-course exam routine (Wenzel, 2010). While a review of the CPS policies (at http://policies.cps.k12.il.us/) reveals no hard-and-fast rules about end-of-course exams, communications with principals informed me that indeed, local schools make policies about the relative importance of end-of-semester exams in their classrooms (Mather, 2014). A review of local news media (Farmer & Selch, 2010) and blog posts (Salazar, 2013) finds some articles advocating for fewer and less significant final exams, evidence that end-of-semester and course final exams are a “big deal” for CPS teachers. Emerging validation studies show the utility of classroom artifacts to “illuminate features of instruction not apparent even through direct classroom observation" (Martínez, Borko, & Stecher, 2011) which given the centrality of the high school end of semester exam, clearly make this an aspect of high school teaching and learning worth studying.

High school end of semester exams are an important artifact of practice. Since these exams are all generated by teachers, and have high accountability for students, they provide an insightful lens into the decisions teachers make about the content and skills they think are most important for students.

Assessment And Instruction Are Tightly Connected

Second, it is clear that assessment and instruction are tightly connected. John Dewey made this connection back in 1916: when discussing developing classroom instruction, he commented how the "aim" or goal of instruction influences all that comes before: "the aim as a foreseen end gives direction to the activity... it influences the steps taken to reach the end" (Dewey, 1916). Moreover, Dewey also promotes a notion of curriculum as experience; that is, if students experience it, then it is part of the curriculum of school, and therefore one central connection between assessment and instruction exists *de facto*.

> *"Hence to develop and train [the] mind is to provide an environment which induces such activity. On the other side, it protects us from the notion that subject matter on its side is something isolated and independent. It shows that subject matter of learning is identical with all the objects, ideas, and principles which enter as resources or obstacles into the continuous intentional pursuit of a course of action. " (Dewey, 1916)*

Dewey argues that the subject matter is not isolated and independent, but part of the overall experience (the "objects, ideas, and principles") of the learner. Assessments are part of instruction. Later philosophers of education have connected Dewey's ideas to formative assessment (Kucey & Parsons, 2012) and professional development (Desimone, Porter, Garet, Yoon, & Birman, 2002). Since philosophically assessment and instruction are so interconnected, even inseparable as part of the continuum of instructional experience, using teacher generated final exams as strong evidence of instruction makes sense; exams are a part of that instruction. And as the next point emphasizes, exams are a distinctively important part.

High School Exams Are Important Tasks, And Tasks Predict Performance

It is a simple fact of schooling that students take tests seriously. As Doyle mentions,

> *"Students tend to take seriously only that work for which they are held accountable. If no answers are required or any answer is accepted, then few students will actually attend to the content." (Doyle, 1983)*

But exams are more than just something that students worry about: they are a particularly powerful and common academic task. Doyle defines academic tasks as:

> *Academic tasks are "(a) the products students are to formulate, such as an original essay or answers to a set of questions; (b) the operations that are used to generate the product, such as memorizing a list of words or classifying examples of a concept; and (c) the 'givens' or resources available to students while they are generating a product, such as a model of a finished essay supplied by a teacher or a fellow student." (Doyle, 1983)*

High school end of semester exams clearly fit this description. Because end of semester exams occur in most high school courses and are largely summative, the nature of the academic tasks that they are comprised of can be a powerful indicator of the learning demands of the whole course. Richard Elmore uses Doyle's definition as the basis for his construct of the "instructional core," which is, loosely, the interaction between teachers, students, and content (City, Elmore, Fiarman, & Teitel, 2009). Elmore's "first principle," that "increases in student learning occur only as a consequence of improvements in the level of content, teachers' knowledge and skill, and student engagement," serves as a powerful lens when analyzing reform efforts.

Additionally, task predicts performance. Again, from Elmore:

> *"What determines what students know and are able to do is not what the curriculum says they are supposed to do, or even what the teacher thinks he or she is asking students to do. What predicts performance is what students are actually doing. Memorization tasks produce fluency in memorization and recall, not necessarily understanding. Memorizing the elements of the periodic table is not the same as understanding the properties of the elements." (City et al., 2009)*

Certainly, high school end-of-semester exams fit these criteria—they are clear depictions of what students actually must do, with considerable stakes attached. Since task predicts performance, an analysis of those tasks will help illuminate important aspects of teacher and student performance.

Teachers Are Taught To Align Their Assessment To Instruction

Fourth, there is evidence that teachers are taught to align their instruction and assessment to standards. In the Chicago Public Schools *Design For High Schools* from 1997, a key priority is "[teacher] learning opportunities... to [align] curriculum and instruction with standards and assessments" (Schools, 1997). Moreover, schools are "required" to use "the academic standards and frameworks to drive core curriculum and instruction," which as described above, also includes assessment. In 2004, the CPS board policy was changed to require "students have passed a series of academically challenging courses in the core subject disciplines of English, Mathematics, Science and Social Studies… that are aligned to the Illinois Learning Standards" [emphasis in the original] (Chicago Public Schools, 2004). Given that teachers do write their own final exams and are told to and taught to align these exams to standards, an analysis of these exams should provide useful insights into how teachers go about the business of determining content and rigor for their courses.

Exposure To Concepts Implies Opportunity To Learn

There is increasing evidence that connects students' opportunity to learn with access to content. When exploring this issue with the 2012 PISA exam, Schmidt and colleagues comment:

> *Many international comparisons of education over the past 50 years have included some measure of students' opportunity to learn (OTL) in their schooling. Results have typically confirmed the common sense notion that a student's exposure in school to the assessed concepts, operationalized in some sort of time metric, is related to what the student has learned as measured by the assessment. (W. H. Schmidt et al., 2014)*

Moreover, there is increasing evidence that thanks to a reasonably coherent worldview of mathematics teaching and learning starting from the first NCTM standards (Usiskin, 2013), the alignment between standards, assessments, and accountability has contributed to the larger improvement in mathematics NAEP scores when compared to reading (Willingham & Grissmer, 2011). Building on these ideas, it is quite reasonable to expect that an analysis of teacher-generated assessments and the nature of their alignment to other assessments, standards and instructional resources will tell us something useful about the opportunity to learn in Chicago science classrooms and the ability of the system to collectively improve.

There Is No Good Reason To Reject The Premise That Final Exams Provide Such Evidence

Finally, the chief arguments against such a framework do not hold up well to scrutiny.

Consider one possible alternative: science teachers do not value the final exam because the most important assessment takes place formatively, throughout the course, so they construct final exams without expecting them to demonstrate what has been taught and learned. If this were true, we would certainly hear about it in the research about assessment practices within the district. In the largest formative assessment pilot done in the district over the past ten years—albeit focused on mathematics (Bogaert, Fendt, & Vogele, 2006)—there is no reference of it. In the history of science improvement efforts in Chicago (Wenzel, 2010), there is no description of this either. It is likely the district wouldn't spend considerable resources on assessment tools (see http://www.cps.edu/SchoolData/Pages/Assessment.aspx#sub2-tab) if classroom teachers had the formative data they already needed on hand.

Consider a second possible alternative: teachers are forced by district policy to use an anemic exam that has no consequence for grading or learning and bears no relationship to teaching and learning. As mentioned above, there is no CPS policy about final exams, but a clear directive that course grades matter for students in the high school promotion policy. CPS teachers are taught and told to align their instruction and courses with standards. But as we will see in Chapter 4, it was relatively easy to collect final exams from many classrooms, and not only are the all different, but they are different within schools for the same course. That would not be the case if district policies were more prescriptive.

A third alternative is that teachers are incredibly inept at creating final exams, so the ones they create bear only a marginal relationship to classroom instruction: in effect, that teachers teach one thing but test another. Again, there is little or no evidence to support this conjecture. If this were true, given the attention final exams have as a routine in the system, we would have heard about by now. I could not find any academic or mainstream literature decrying a final exam problem in Chicago—at the very least, students and parents would be pretty upset if a high stakes final exam didn't match the course content. There was one pro-and-con article in the Chicago Tribune about final exams (Farmer &

Selch, 2010)—indicating there might be some debate about their overall utility, and one blog post (Salazar, 2013) talks about final exams as a problem more for the grading implications than the assessment ones. Neither of these articles is about science, and neither is substantive enough to think that teachers don't take the final exam routine seriously.

In sum, there just is not any good evidence that teachers do **not** take final exams seriously as indicators of what has been taught and learned. Whether they are adept at creating them or not, teachers expect final exams to tell us who learned what in the course of a semester. Like homework, final exams are a part of what Stigler and Hiebert refer to as the "culture of American teaching" (Stigler & Hiebert, 1999) and as my focus-group research confirms, teachers tend to believe that final exams matter.

Conceptual Framework

As mentioned in chapter 1, I will use teacher-generated end-of-semester exams as actual artifacts of individual classroom practice. In doing so, I will be able to explore an important part of the technical core of high school science teaching and learning—both the specific routine of "the final exam" as it exists in Chicago high school science classrooms, but also the degree to which those final exams do or do not have any connection to the instructional practices of teachers. I expect to create a picture of classroom practice that explains, based on these documents, what high school science teachers in Chicago value about the content they teach, how they organize that content, and the sort of demands they expect of their students. In the discussion above, I showed that the literature base does not currently answer these questions. I also demonstrated that we can learn something important about science instruction by analyzing final exams—or more pointedly, that final exams provide important evidence about the instructional core. I will now specific the research questions that will drive this endeavor.

Research Questions

As Hatch mentions, "identifying research questions is a critical step in research design because questions give direction to the study, limit the scope of the investigation, and provide a device for evaluating progress and satisfactory completion" (Hatch, 2002, page 41). Using a large collection of high school science teacher generated end of semester exams, I have explored what is happening in high school science classrooms in the Chicago Public Schools, determining the depth and breadth of science content in Chicago public high schools, and how classroom teachers measure it. This is framed by the following three questions.

1. **Content**: What content is addressed by Chicago high school science teachers in the courses they teach, as reflected in the assessments they use? How does this content relate to state and national standards? How are topics with important reform or societal connotations (*e.g.* evolution, the environment, inquiry, nature of science) treated? Finally, does the content indicate anything about what the teachers value, or are there other explanations for why some content is included and other content not?

2. **Rigor**: What is the level of "academic press" associated with Chicago high school science courses, as reflected in the assessments used in the courses they teach? How does this level vary between schools?

3. **Assessment Literacy**: What does local assessment design, particularly on final exams, say about the level of assessment literacy of CPS high school science teachers? Are there explanations other than "assessment literacy" that might explain trends in design of these final exams?

Theoretical Perspective

This research creates a picture of practice at the classroom level. As such, it is certainly much more theory development and hypothesis generation than theory application or hypothesis testing. The practices of determining curriculum, selecting instructional materials, identifying topics, sequencing concepts, designing and delivering instruction, addressing rigor, and then designing and administering assessments constitutes an opaque process in most high school science classrooms (Weick, 1976). The bulk of my research work consists of collecting and analyzing data in the form of teacher-generated final exams that describes in very concrete terms the instructional decisions that are made by teachers and teams of teachers; from this, I will then craft some hypotheses that might explain the sense-making processes teachers and school use, with implications to leadership, school organization, and policy.

However, late 2013 is an exciting time for science education and positioning this research within the history of science education is important. In 2012, the National Academy of Science released the *Framework For K-12 Science: Practices, Crosscutting Concepts, and Core Ideas*, which defines the science content American students should know (National Research Council, 2012). Last year, Achieve, Inc. released the *Next Generation Science Standards* (NGSS), a set of content standards developed by states based on the *Framework*

(NGSS Lead States, 2013a). These new standards have already been adopted by six states, with others—including Illinois—ready to adopt them soon. They represent a dramatic change in vision and substance from the existing Illinois standards (Jackson et al., 2013) and an unprecedented event in the history of science education in the United States (DeBoer, 1991). We do not have many reasons to believe that instructional practice has shifted dramatically in the past decade in science classrooms since the previous set of Illinois science standards, but to reach the goals of the NGSS will clearly require dramatic change. In this context and history, this study serves to set a baseline that describes current science teaching and learning practice. As the new standards are adopted—with all the curriculum, assessments, and supports that are implied—we should hope to see changes in instructional practice that lead to more student learning. This research seeks to empirically define the starting point for those efforts. We should hope that a similar study, conducted say 5 years after NGSS implementation begins in Chicago, would hope to highlight different results.

<u>The 10-Year Old Data Problem</u>

As mentioned in the first chapter, the set of teacher generated end-of-semester final exams I will use to answer these questions is from the 2003-04 and 2004-05 school years. This is data that is ten years ago. Can we expect data this old to tell us anything of value? I believe we can, for the following reasons.

First, we can use this to benchmark a point in time. This data will describe classroom and teacher practice in the past as a benchmark. Subsequent changes in instructional practice and schools could or should be expected to show changes in the design and construction of teacher generated end-of-course exams, and when that study is done the results could be compared to this data. Logistically and practically, there is no way to retroactively collect this information, so using what we have now makes good sense.

Second, with this study I am in part confirming the approach. As I will argue in chapter 6, the sampling and analysis methodology that I am using here to make claims about classroom practice and instruction does not depend on the school year or the amount of time that has passed, and returns useful insights about schools and classrooms in a fashion that is both trustworthy and affordable. The approach is worth exploring, regardless of the age of the data.

My Background

My experiences as a curriculum developer and teacher have certainly colored this work. As a former science teacher who has studied and developed curriculum using ideas about conceptual flow, learning progressions, and a sensibility about misconceptions (The Geographical Data in Education (GEODE) Initiative, 2005), my analytic lens certainly been shaped by those sources. As a district leader who has used and championed instruction-based reforms to improve student learning, understanding how teachers and leaders collectively make sense about the content that is taught and the instruction that is subsequently delivered is front-and-center to my work. My generally progressive viewpoint, meshed with a firm grounding in science content and epistemology, will certainly bring my focus back again to the content at hand.

And my own background as a district leader clearly complicates things. I had an integral role in leading science education improvements in the Chicago Public Schools from 2002 to 2009. As leader, my charge was to champion and drive the sorts of policies and programs that will enhance student achievement at all levels, and the work that has resulted from the exam repository to some degree comes from my own vision. While there is not an ethical conflict with pursuing these lines of inquiry, surfacing my role as the district's curriculum leader in this case is essential. One of my intentions in pursuing this line of study is to learn about the practice of teaching and learning science in a manner that informs the practice of district leadership.

Essentially, then, this becomes my approach: Little information exists that describes the technical core of high school science teaching and learning, apart from broad high level evaluation documents that describe poor student performance based on large scale assessments or scholarly treatises based on hothouse examples that tell policy-makers what to do. Through this research, I have attempted to generate something in between, which collects many artifacts of instructional practice, synthesizes meaning among and between them, and then connects that analysis to system-level data about organization and performance, and identifies the at-scale implications.

Since I have taught high school science for nine years, and been involved in district science leadership for several more, issues that connect teaching and learning with school leadership are close to my heart. There is an increasing demand to improve mathematics and science education throughout the United States (Members of the 2005 "Rising Above the Gathering Storm" Committee, 2010), with issues of science, technology, engineering, and mathematics (STEM) education receiving direct attention from President Obama (The White House, 2012), yet the research base about what is actually happening at the classroom level is remarkably thin. While standardized test scores provide a measure of achievement outcomes, a more intense look at what is taught in high school science classrooms, using the lens of what teachers actually assign and assess, will give school leaders, district administrators, and policymakers a more complete picture of the current state and implications for change.

Chapter Conclusion

In this chapter, I identified the gaps in the literature base that was surveyed in chapter 2. I posited an approach that both helps to fill the gaps in the research base and is based on an analysis of teacher-generated end-of-course final exams. Based on this information, I articulated my research questions, explained how my background frames my understanding of them, and discussed the data that will be used to generate findings. In the next chapter, I will discuss the methods I used to generate answers to these questions.

Chapter 4: Methodology

Introduction

In this chapter, I will describe the methods used to collect and analyze the data to answer the research questions described in the previous chapter.

Methods Overview

A summary of the sources of evidence, the particular measures, and the research activities used by research question are displayed on Table 1. The primary evidence source used to generate these answers is many teacher-generated end of semester exams, plus a focus group discussion of Chicago Public Schools (CPS) high school science teachers. Measures for content and rigor came from the chosen analysis tool, the *Survey of Enacted Curriculum* (Rolf K. Blank, Porter, & Smithson, 2001), as described below; measures for assessment literacy came from an anaylsis of item type distribution and from focus group comments. There are three primary research activity types involved: document collection and review, frequency counting (and associated statistics) based on the data set of assessments and items, and analyzing the focus group transcrips.

Research Questions		Evidence	Measure	Research Activity
1	Content	Teacher-generated high school science semester exams collected from 2003-05; Focus group of CPS high school science teachers	Item content scores on the *Survey Of Enacted Curriculum*	Document review Frequency counting Transcript analysis
2	Rigor		Item cognitive demand scores on the *Survey of Enacted Curriculum*	Document review Frequency counting Transcript analysis
3	Assessment Literacy		Item type and distribution, Focus group comments	Transcript analysis

Table 1: Evidence, Measures, and Research Activities by Research Question

The Recipe

While details follow in the sections below, what follows is the step-by-step process used to collect this data and conduct the research.

1. **Collect exams**. As part of my work for the Chicago Math & Science Initiative, I collected end-of-semester exams from every Chicago high school that we could. These exams were scanned to PDFs, labeled with a school identifier number, and organized into an electronic database.
2. **Identify schools**. Using the publically released data available for the Chicago Public Schools, I generated lists of all high schools in the district, including demographic and performance data. This information was added to the electronic database.

3. **Organize school and exam data tables**. Since the exams spanned multiple schools, courses, semesters, and school years, a relational database was created to track and organize all this information. Each exam has multiple questions, adding an additional dimension to the data.
4. **Code exams for content and intellectual demand**. Using a well-established protocol and a team of trained reviewers, each question on each exam was coded for item type, content, and intellectual demand. Since discerning this information involved some professional judgment on the part of the reviewer, a mechanism was established to maximize inter-rater reliability and minimize time to perform this work.
5. **Test findings in a teacher focus group**. I convened a focus group of Chicago high school science teachers to discuss the nature of high school science exams in Chicago and to get their perspective on the findings that were emerging from the data.
6. **Generate findings**. Queries of the database are used to generate findings. This process is explained in more detail in the next chapter; essentially, fields can be sorted and filtered to compare items (and proportions of items) by course type, school type, and other variables. These findings are then matched and connected to themes that emerged from the focus group.
7. **Turn findings into analysis**. In the last chapter, I will take the list of findings and explain what they mean, in light of both school practice and the current literature. I will explain what various levels of leadership can and should do to address some of these findings, the implications for practice at the school and district level, and implications for future research.

Organizing Data

Since this research involves many exams across many schools, each with many questions, organizing this "document review" information is a major challenge. The end products of the document review work were three related database tables. The three data tables are linked to one another and are analyzed mostly via frequency counts and proportions. When coupled with the focus group transcript, this presents a robust and unified source of data from which to generate findings.

The first data table, called the "school table," describes school characteristics based on performance and demographics. This includes information about the school's performance, location, demographic make-up, and similar characteristics, all pulled from official Chicago Public School sources.

The second, called "exam table," describes each of the exams that I collected. Each exam was identified by course name, by school year, by semester, and by school. By combining the school table and the exam table, I can ensure that I have an appropriate mix of content, geography, and school year in my sample of data. The exam table also includes data about the exam as a whole, such as the number of items on the exam.

The third table, called the "item table," contains item level data for every exam in the study. This table includes a description of every exam question, the content assessed, and the level of cognitive demand a correct answer requires. Since making these determinations involves some degree of subjectivity, to generate this table I needed to devise a process to engage multiple reviewers of each item and verify that the coding process was accurate.

A Note About Technology

In general, these data tables and analysis were performed using Microsoft Excel 2010 and 2011 and storied on a personal Microsoft SharePoint 2007 site. Linking between them was done using built-in Excel functions, and data verification and cleaning was also done via a variety of text-scrubbing tools and Excel macros. Basic statistical analysis was done on Microsoft Excel as well, with a few tests run in SPSS.

The School Table

Generating a table that describes every high school in Chicago is not particularly difficult work. On September 25, 2009, extensive data tables describing school and school performance were downloaded from the Research and Evaluation website of the Chicago Public Schools. Some straightforward cutting and pasting enabled me to create a table that sorted schools by unit number (the unique school identifier used by the Chicago Public Schools), instructional area (a measure of geography), number of students enrolled, the proportion of students reaching the meets or exceed threshold on the Prairie State Achievement Exam (PSAE), the percentage of students who received free or reduced lunch. Since I am focusing on the 2003-04 and 2004-05 school years, to make a master data table I included data from both years, and wrote code to manage the exceptions. (For instance, North Grand HS was a new school at the time and did not have PSAE results in 2004.)

It is a common routine within the Chicago Public Schools to change the instructional region or area ("area") designations of schools. To ensure geographic accuracy, I used the latest numbers at the time of download—area designation for the 2009-10 school year. Some of the schools in my study, based on 2003-05 data, have been closed or changed in the intervening time, and are assigned correspondingly new area numbers. Using the most up-to-date area assignments ensured that links between tables in my data would be valid.

In order to make analysis easier, I also created some categorical data of key variables. Table 2 shows school size categories. Table 3 shows school performance categories, based on the PSAE Science scores. And Table 4 shows school poverty categories, based on the percent of students receiving free or reduced lunch. All data from here come from the same Chicago Public Schools data set.

Enrollment	**School Size**
Less than 500	"small"
Between 500 and 2000	"medium"
More than 2000	"large"

Table 2: School Size Categories

Percent Of Students Meeting or Exceeding on PSAE Science in 2003-04	**Performance category**
Less than 25%	"low"
Between 25% and 50%	"medium"
More than 50%	"high"

Table 3: School Performance Categories

Percent Of Students Receiving Free or Reduced Lunch	**Poverty category**
Less than 60%	"low poverty"
More than 60%	"high poverty"

Table 4: School Poverty Categories

The Exam Table

The exam table is a list that describes the key characteristics of each exam. Each exam is linked to a particular school in the school table, but also has information unique to that exam.

Collecting The Exams

The exams were collected in the 2003-04 and 2004-05 school year as part of the high school improvement efforts of the Chicago Math & Science Initiative. Each school was asked to participate in a 6-step process designed to accelerate the development of a collaborative team-based structure within the mathematics and science departments. The initial request from December 2003 is shown on Figure 1, from a slide presentation delivered at a fall meeting of all high school science department chairpersons. Collection and sharing of exams was "phase 4" of this work.

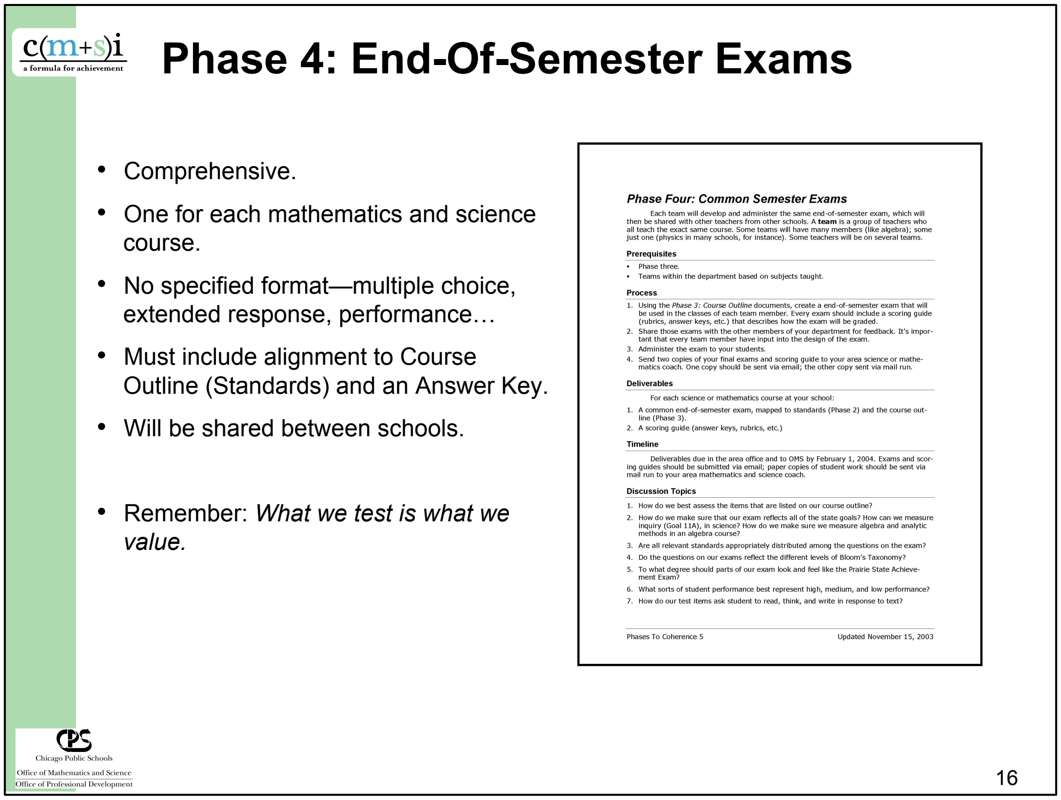


Figure 1: Initial request for common end-of-semester exams from CPS Science department chairs, December 2003

As exams were sent in to the Chicago Public Schools Office of Mathematics and Science, a team of interns scanned each document to a PDF and uploaded them to a SharePoint site to collect and track them. For each exam, they identified the school unit number, school year, course name, and semester (*e.g.* first semester or second semester) by fields on SharePoint. As the original plan was to have this information fed back to teachers and schools, we crosschecked each exam with multiple reviews, making sure the school and course were recorded correctly. These exams (and an accompanying website) were used periodically for small-group professional development sessions over the next few years, providing an ample opportunity to correct any transcription errors that might have arisen in the scanning and collecting process.

For my research, a new SharePoint site was created on a personal website, and the directories and contents were copied directly from CPS to this site, preserving all the essential data and metadata.

Managing The Exam Table

I created several additional fields to enable the analysis of each of the exams. Fields were created to track up to three individual reviewers for each exam. (Reviewers are the people who will conduct the coding of the exam to determine the content and cognitive demand of the items.) I also included fields to track the status of each review—has the exam been shipped to the reviewer, had it been returned, had the individual item codes been entered into the computer, etc. I also created an anonymous key code that could be shared with reviewers, so that the name of the school in question would remain anonymous.

Selecting Exams To Code

I knew that coding exams would be a time consuming and expensive task, so I decided to limit the number while still enabling me to answer my research questions. The original collection of exams is shown on Table 5, organized by school year.

	Anatomy	**Astronomy**	**Biology**	**Botany**	**Chemistry**	**Earth and Space**	**Environmental**	**General Science**	**Physical Science**	**Physics**	**Grand Total**

	Anatomy	Astronomy	Biology	Botany	Chemistry	Earth and Space	Environmental	General Science	Physical Science	Physics	Grand Total
2003-04 All	3	3	46	1	62	37	43	2	0	37	234
2003-04 Semester 1	1	2	32	1	44	26	33	2	0	29	170
2003-04 Semester 2	2	1	14	0	18	11	10	0	0	8	64
2004-05 All	2	1	37	0	15	8	11	0	1	9	84
2004-05 Semester 1	2	1	34	0	15	8	11	0	1	9	81
2004-05 Semester 2	0	0	3	0	0	0	0	0	0	0	3
Grand Total	5	4	83	1	77	45	54	2	1	46	318

Table 5: Initial Exams Collected, By School Year

Clearly, coding 318 exams would be both cost and time prohibitive. I wanted a smaller collection that would still enable me to provide a strong perspective about science teaching and learning in the Chicago Public Schools.

To get as broad a perspective as possible, I decided to code all the biology exams. Since this was the course with the most data, this would provide me with as robust a district-wide perspective as possible.

But I wanted to focus on science instruction, not biology instruction. For further analysis, I quickly eliminated exams from the minor courses: anatomy, astronomy, botany, general science, and physical science. Chemistry was easy to pick, as they represented the greatest quantity of remaining exams. I decided to also focus on physics instead of earth and space science or environmental science, for the following reasons.

- The sense of what constitutes a typical physics curriculum content quite strongly defined (Arons, 1990), likely more so than in the newer field of environmental science.

- The biology, chemistry, physics sequence is a common one in the Chicago Public Schools (LaForce & Feranchak, 2003).

- According to ACT, the biology-chemistry-physics sequence is the one that corresponds with highest ACT science scores (ACT Inc. & The Education Trust, 2004).

In order to make the coding manageable, I randomly selected a mixture of chemistry and physics exams to review, sampling the available dataset so as to be able to make reasonable predictions about the content and intellectual demand of those courses. The initial set of "active" exams is shown on Table 6. As will be seen in the subsequent section, not all of these exams had valid items to be coded, so some were ultimately eliminated from consideration from the overall analysis.

	Number of Active Exams			
	Biology	**Chemistry**	**Physics**	**Grand Total**
2003-04	46	8	11	65
Semester 1	32	3	8	43

	Number of Active Exams			
	Biology	**Chemistry**	**Physics**	**Grand Total**
Semester 2	14	5	3	22
2004-05	37	3	4	44
Semester 1	34	3	4	41
Semester 2	3	0	0	3
Grand Total	83	11	15	109

Table 6: Initial Selection of "Active" Exams

The Item Table

The heart of my analysis occurs on the item table, which is a large array with information that describes each item on each exam in the study. First, I needed a scoring protocol or tool that would enable me collect the data I needed; then I needed to develop a process to have multiple reviewers code each exam, and finally, I needed a mechanism to assemble all the data. I will address each of those steps in turn.

Review Protocol: The Survey Of Enacted Curriculum

Several possible protocols can be used to align items to content and intellectual demand—any number of individual state level tools, to the Project 2061 benchmark analysis tools, to the various frameworks utilized by large scale exams like NAEP and PISA. But it was clear that the *Survey of Enacted Curriculum* (Blank, Porter, & Smithson, 2001) made best sense as the right tool to use for this purpose. There are several reasons for this.

First, the principal portion of the *Survey of Enacted Curriculum* tool addresses both cognitive demand and content alignment in a manner that can be applied to any high school science course. This is important, as different frameworks or tools for different courses would have made the data collection even more challenging. Second, since the Council of Chief State School Officers (CCSSO) contributed to the genesis of the *Survey of Enacted Curriculum*, it has considerable momentum in the field. The large NSF Math-Science Partnership grant that CCSSO received was based in part on implementing *Survey of Enacted Curriculum* in many states (Blank et al., 2005), and it was a required component of the Illinois Math-Science Partnership grants in 2004 and 2005 (RMC Research Corporation, 2005). Third, the alignment index on the *Survey of Enacted Curriculum* report shows a high correlation with impact on student achievement (Gamoran, Porter, Smithson, & White, 1997). And finally, it was easy to find reviewers who were knowledgeable of the *Survey of Enacted Curriculum* to help me with my analysis. While the *Survey of Enacted Curriculum* is not aligned to the Next Generation Science Standards (NGSS Lead States, 2013b), there are not any existing tools that do so and the exams in this data set precede NGSS by several years.

One drawback with the *Survey of Enacted Curriculum* is that it fundamentally is a reductionist tool: by design, it provides for the analysis of discrete items on an instructional artifact, and does not look at the connections and interdependencies between items. Recent reports, such as *Systems for State Science Assessment* (Committee on Test Design for K-12 Science Achievement, 2005) and *Developing Assessments For Next Generation Science Standards* (Committee on Developing Assessments of Science Proficiency in K-12, 2013) both discuss assessments as part of coherent systems with many interlocking pieces. However, as will be shown in Chapter 5, the relatively low level of cognitive demand, broad swath of content coverage, and lack of item type diversity on these exams indicates a relatively unsophisticated design, while probably means there are not many connections or interdependencies to find among items. Additionally, this study intentionally focuses on district-wide scale and samples pieces and aspects of exams to paint a picture of the district whole, making a classroom-level analysis of item relationships less important. That said, I will revisit the choice of the *Survey of Enacted Curriculum* in the final analysis.

Details about the *Survey of Enacted Curriculum* coding procedure used are online at http://seconline.wceruw.org/Reference/CodingProcedures2008.pdf. To code an item using the *Survey of Enacted Curriculum* process for this project, reviewers had to record data for the following fields for each test question ("item").

Item Number

The number of the item on the exam. The first question on the exam would be item number one, the second item number two, etc.

Author Item

The number of the item according to the test author (the teacher). Frequently, teacher-written exam questions contain multiple questions within one numbered question—for instance, a fill-in-the-blank sentence listed as one numbered question, but with two blanks for two possible student responses.

Item Format

Reviewers determined the format of the item. Possible item formats are listed on Table 7.

Item Format	**Notes**
Constructed response	Any question demanding more writing than a "short answer" question, typically at least a full sentence answer, and sometimes, many paragraphs.
Fill in the blank	Students are asked to identify a missing word or phrase from a given sentence fragment.
Multiple-choice	Students are asked to choose the correct response from a number of choices given. Questions that were matching (*e.g.* match the label on the diagram of a cell with the proper scientific term) were also considered multiple choice. Reviewers recorded the number of possible options for each item as well (*e.g.* "multiple-choice – 4" or "multiple-choice – 5").
Rewrite sentence	Students were asked to rewrite the given sentence—a slightly more complicated version of fill-in-the-blank.
Short answer	Students were asked to respond with the results of a calculation or a short, 1-3 word answer.
True or false	Students were asked if a statement was true or false.

Table 7: Possible Exam Item Formats

Item Weight

Reviewers were asked to designate "how many points" this particular question was worth on the exam, based on exam instructions or scoring guides provided to the student. If no clear item weight was given, the default value of one was assumed.

Content

Content was coded using the *Survey of Enacted Curriculum* coding categories, four-digit numbers that correspond to all the content generally possible in high school science courses. Since sometimes it is hard to determine where "molecular motion" ends (code 1801) and "pressure" (code 1802) begins, reviewers can select up to three different content codes for each item. The order of the content code is not relevant. The *Survey of Enacted Curriculum* content coding categories are shown in appendix 1 on page 144. Reviewers coded exams using the most fine-grained categories possible.

Cognitive Demand

The *Survey of Enacted Curriculum* allows for five different levels of cognitive demand: memorize, perform procedures, communicate understanding, analyze information, and apply concepts/make connections. These are show on Figure 2. One clarification was made to the categories: Questions that call for students to look information up on a chart, table, or graph—including interpolating and extrapolating—were considered "perform procedures," so we added that as part of our category descriptions.

Cognitive Demand Categories for Science

B	C	D	E	F
Memorize	**Perform Procedures**	**Communicate Understanding**	**Analyze Information**	**Apply Concepts / Make Connections**
Recite basic science facts	Make observations Collect and record data	Explain concepts	Classify and compare data	Apply and adapt science information to real-world situations
Recall science terms and definitions	Use appropriate tools Make measurements, do computations	Observe & explain teacher/student demonstrations	Analyze data, recognize patterns	Apply science ideas outside the context of science
Recall scientific formula	Organize and display data in tables or charts	Explain procedures & methods of science & inquiry	Reason inductively or deductively	Build or revise theory/plan and design experiments
	Execute procedures Conduct experiments	Organize & display data in tables or charts	Draw conclusions	Synthesize content and ideas from several sources
	Generate questions, make predictions	Student presentations of science information	Identify faulty arguments or misrepresentations of data	Use and integrate science concepts
	Test effects of different variables			

Figure 2: Cognitive Demand Categories for Science

Coordinating Reviewers For The Item Table

Loyola University's Center for Science & Math Education (online at http://www.luc.edu/cse/index.shtml) had a team of people trained in conducting *Survey of Enacted Curriculum* analysis of high school exams. After several calls with Dr. John Smithson of the Wisconsin Center for Education Research, and the project director/lead author of the *Survey of Enacted Curriculum* project, he agreed that this was the best and most effective strategy for this research. Training my own set of reviewers would be possible but difficult; and hiring WCER to do the review work would be considerably more expensive that the local Loyola team.

The Loyola team consisted of former high school science teachers who had all participated in the WCER-led training sessions on the *Survey of Enacted Curriculum*. They had also used the *Survey of Enacted Curriculum* as part of their support work with schools and districts, and so were quite experienced users of this exam coding tool.

An agreement was reached with the leaders of the Loyola University Center for Science & Math Education to have each of the 109 active exams read and coded by two of their staff members over the spring and summer of 2010. I sent exams in batches as scanned PDF files to Loyola, identified only by the subject and an anonymous key. The returned a series of Excel sheets with the consensus reviews from their reviewing team. That last point is an important one—by having each exam reviewed by two people, and forcing those two reviewers to discuss their coding decisions and reach a consensus decision, I ensured the quality of their ratings was high. Since all communication was electronic, transcription errors were minimized.

During the review process, the Loyola team found 14 of the 109 exams to be incomplete or illegible in some manner. Upon inspection, these were exams with missing pages or illegible scans, and so they were rejected from the analysis.

Verifying The Reviews

Despite the strong background, training, and experience of the reviewers, I wanted to make sure that the review process was accurate. I decided to perform two checks on the data to make sure it was valid—to check that results were both internally and externally valid.

First, I wanted ensure the work was internally valid: that the consensus process used by each reviewer pair was working. If the reviewers were generally in agreement, and had to only come to consensus occasionally, that would indicate that it was relatively straightforward for them to come to decisions about item content and cognitive demand, and their judgment likely represented the "true" value. On the other hand, if the reviewers disagreed frequently, and had to spend lots of time and energy coming to consensus, that would indicate that their preparation was not very strong or the coding rubrics and tools were not sufficient to clearly distinguish between item content and cognitive demand levels.

To check this, I asked the Loyola team to initially provide the coding sheets from 20 exams and 40 reviews without the consensus process. This represented 1044 individual items and all individual reviewers. As is shown on Table 8, the number of disagreements between reviewers is quite low. This indicates that the consensus dialogue did not need to occur very often because the initial ratings of the reviewers were so often the same, giving me confidence that the Loyola team has been well trained to execute the *Survey of Enacted Curriculum* coding procedure.

Variable	**Number Of Disagreements**
Item Type	6 of 1044 items
Content	34 of 1044 items
Cognitive Demand	21 of 1044 items

Table 8: Determining Internal Validity On Loyola Exam Coding

I then needed to check to make sure that the results of the Loyola were externally valid; namely, would another, external reviewer come up with a similar decision if they followed the same process. To gauge this, I scored 39 exams myself, and then compared my scores with those of the Loyola consensus team. The results are shown on Table 9, and from this data, it is clear that there were very few disagreements. This means that my own judgment matches quite strongly with the consensus of the Loyola reviewers, and gives me quite a great deal of confidence that the results of the coding process would be deemed valid to other external reviewers.

Variable	**Number Of Disagreements**
Item Type	0 of 1989 items
Content	10 of 1989 items
Cognitive Demand	2 of 1989 items

Table 9: Determining External Validity Of Loyola Exam Coding

Because the high level of external validity of the coding process, I decided that it would be acceptable to only use one reviewer for the Chemistry exams—and certainly faster and less expensive. While this does not provide assurance that the chemistry content is externally valid, given the high degree of agreement in other subjects it is reasonable to presume the Chemistry results are not very different. For all the biology and physics exams, more than one reviewer was used to generate the *Survey of Enacted Curriculum* codes.

Assembling The Data For The Item Table

Once the exams had been coded, and the coding process verified to be both internally and externally valid, the data was combined into one master data table for analysis. This was a data set consisting of 95 semester and final exams encompassing 4776 items.

On Table 10, distribution of items by instructional area is shown. These are the 2009-10 area classifications, a grouping of schools into roughly geographic boundaries that changes slightly each year. On Table 11, the number of exam items is shown by school size. Taken together, these tables indicate that the breadth of items generally spans the geography and school type in Chicago.

Area	**Number Of Items**			
	Biology	**Chemistry**	**Physics**	**Grand Total**
19	631	116	0	747
20	25	0	0	25
21	446	64	97	607
22	50	0	0	50
23	617	50	189	856
24	497	0	36	533
25	65	0	15	80
26	642	163	13	818
28	258	0	0	258
29	49	0	0	49
54	496	185	72	753
Grand Total	**3776**	**578**	**422**	**4776**

Table 10: Distribution Of Items By Instructional Area

School Size	**Total Number Of Items**			
	Biology	Chemistry	Physics	Grand Total
large	1036	284	90	1410
medium	2204	294	251	2749
small	536	0	81	617
Grand Total	3776	578	422	4776

Table 11: Distribution of Items By School Size

On Table 12 and Table 13, the number of items by semester and school year are shown.

Semester	Biology	Chemistry	Physics	Grand Total
1	2977	326	409	3712
2	799	252	13	1064
Grand Total	3776	578	422	4776

Table 12: Distribution Of Items By Semester

School Year	Biology	Chemistry	Physics	Grand Total
2003-04	2292	418	287	2997
2004-05	1484	160	135	1779
Grand Total	3776	578	422	4776

Table 13: Distribution Of Items By School Year

On Table 14 and Table 15 the distribution of items by poverty level and performance level are shown.

Poverty Level	Biology	Chemistry	Physics	Grand Total
high poverty	3342	478	347	4167
low poverty	434	100	75	609
Grand Total	3776	578	422	4776

Table 14: Distribution Of Items By Poverty Level

Performance Level	Biology	Chemistry	Physics	Grand Total
high	162	100	39	301
low	2212	269	295	2776
middle	1082	209	88	1379
no data	320	0	0	320
Grand Total	3776	578	422	4776

Table 15: Distribution Of Items By Performance Level

Is This A Representative Sample Of Biology Exams?

The final set of 95 exams (and 4776 items) is shown on Table 16. I now will confirm that the schools represented here is a valid sample of the entire district. Because items exist on exams and exams are given by teachers in particular schools, item distribution is a function of school distribution. Since I do not have an exam for every subject, school, semester or year, I need to determine the nature of the claims I will be able to make using the sample of exams I do have. The strategy I will use is based on the recommendations via email conversations with Dr. Evan Smith and Dr. George Karabatsos of the University of Illinois at Chicago.

School Year	Number of Active Exams			
	Biology	Chemistry	Physics	Grand Total
2003-04	41	8	8	57
Semester 1	29	3	7	39
Semester 2	12	5	1	18
2004-05	32	3	3	38
Semester 1	29	3	3	35
Semester 2	3	0	0	3
Grand Total	73	11	11	95

Table 16: Final Selection of "Active" Exams

It is difficult to claim without any doubt this is a representative sample of biology exams that represents the whole district. The process to collect them was not truly random, as any number of factors—like teacher background, principal background, and internal school culture—could have influenced a particular school's decision to send the exams to the central office. That said, some common sense indicates that the current distribution is pretty closely aligned with the overall district distribution of schools. A more formal way to make this argument is to perform a chi-square test where the null proportion is the true population proportion, and if this null hypothesis is not rejected, it would suggest my sample is representative of the whole district. However, issues about randomness would still apply, so performing this calculation is more an exercise.

Therefore, in the discussion below I will do the following.

1. Discuss the distribution of exams
2. Conduct a chi-square test that compares the sample of schools with exams to the sample of the entire school district

Overall Distribution Of Biology Exams

To discuss the distribution of schools with biology exams, I want to compare the sample that I have collected with an overall list of CPS high schools. Unfortunately, obtaining a clean list of CPS high schools is not easy—schools are opened and closed each year, schools have different course taking patterns and might not offer the same instructional program to students, and district classification systems often mean multiple school "unit numbers" occur in the same physical building. I will make some assumptions and clarifications to enable more straightforward analysis.

From data about science course taking patterns in Chicago (LaForce & Feranchak, 2003), I know that while Biology is often a 9th grade course, it is not exclusively so in the Chicago public schools. The Prairie State Achievement Exam is used in Illinois as the NCLB accountability exam, given to 11th grade students. To identify schools that should be included in the comparison set, I will use schools that have a PSAE science score on file. Since I have exams from both 2003-04 and 2004-05, I will use the aligned list of schools that have PSAE scores for either of those two years. Of the 132 high schools listed on the CPS high school data tables, there are 80 high schools with PSAE data. Of the 73 biology exams that are active, only 68 are from schools with a PSAE exam—the others are likely from new schools that have not yet taken the junior year PSAE yet.

From Wenzel's history of mathematics and science reforms in the Chicago Public Schools (Wenzel, 2010), we know there were no major high school science curricular interventions district-wide between 2003-04 and 2004-05 school years. Thus, it is unlikely we would see major systemic differences in end of semester final exams between these two academic years. For this reason, I will group these two years together, including possible schools from either year into the overall list. We also know that at this point there was no set curriculum, instructional materials, or scope and sequence for biology courses in CPS. However, the previous work under the previous administration to produce end-of-course exams ("CASE exams") generated some course definitions (Chicago Public Schools, 2000a), so I will consider four possible sets to compare against the list of all CPS high schools.

1. CPS high schools with a PSAE score and a semester 1 biology exam
2. CPS high schools with a PSAE score and a semester 2 biology exam
3. CPS high schools with a PSAE score and both a semester 1 and semester 2 biology exam
4. CPS high schools with a PSAE score and either a semester 1 or semester 2 biology exam

The distribution of exams is shown on Table 17 (by poverty), Table 18 (by performance level), Table 19 (by performance third), and Table 20 (by size). A few things are clear from a simple review of these proportions.

First, I have collected mostly semester 1 exams. The distribution of these is a close match to the overall distribution of schools. It is reasonable to expect that conclusions drawn from these exams are representative of all schools. Second, the collection of semester 2 exams, while smaller, is still reasonably representative. It is reasonable to expect that the conclusions drawn about semester 2 exams apply are representative of the entire population. Third, it will be difficult to claim that this picture represents all the biology content students are assessed on, as the proportion of schools that have both semester 1 and semester 2 exams is small. In chapter 5, I will compare the content assessed on all semester 2 exams and see if the content is similar across all exams. If all semester 1 exams seem to cover similar content, and all semester 2 exams seem to cover similar content as well, I likely will have a pretty robust picture of content assessed over a student's biology coursework, but because of the lack of schools with both semesters included this argument will not be able to establish conclusively the full range of biology content that students are assessed on. Fourth, it is clear that the claims about representation above apply to distribution of poverty, performance level, performance third, and size. Regardless of the category, the distribution is pretty similar.

CPS High Schools	**High Poverty**	**Low Poverty**	**No Data**	**Total**
Schools with PSAE	71 (89%)	8 (10%)	1 (1%)	80 (100%)
Schools with PSAE and semester 1 Biology exam	43 (90%)	4 (8%)	1 (2%)	48 (100%)
Schools with PSAE and semester 2	11 (73%)	4 (27%)	0 (0%)	15 (100%)

CPS High Schools	High Poverty	Low Poverty	No Data	Total
Biology exam				
Schools with PSAE and both semester 1 and semester 2 Biology exam	3 (100%)	0 (0%)	0 (0%)	3 (100%)
Schools with PSAE and either semester 1 and semester 2 Biology exam	59 (87%)	9 (13%)	0 (0%)	68 (100%)

Table 17: Poverty Category Comparison of Schools with PSAE and Schools with PSAE and Biology Exam On File

CPS High Schools	Performance Level				
	High	Medium	Low	No Data	Total
Schools with PSAE	9 (11%)	13 (16%)	53 (66%)	5 (6%)	80 (100%)
Schools with PSAE and semester 1 Biology exam	6 (13%)	4 (8%)	37 (77%)	1 (4%)	48 (100%)
Schools with PSAE and semester 2 Biology exam	5 (33%)	9 (60%)	1 (7%)	0 (0%)	15 (100%)
Schools with PSAE and both semester 1 and semester 2 Biology exam	0 (0%)	3 (100%)	0 (0%)	0 (0%)	3 (100%)
Schools with PSAE and either semester 1 and semester 2 Biology exam	3 (4%)	20 (29%(	43 (63%)	2 (29%)	68 (100%)

Table 18: Performance Category Comparison of Schools with PSAE and Schools with PSAE and Biology Exam On File

CPS High Schools	Performance Third				Total
	Top Third	Middle Third	Bottom Third	No Data	
Schools with PSAE	29 (36%)	30 (38%)	16 (33%)	5 (6%)	80 (100%)
Schools with PSAE and semester 1 Biology exam	18 (38%)	17 (35%)	11 (23%)	1 (2%)	48 (100%)
Schools with PSAE and semester 2 Biology exam	14 (93%)	0 (0%)	1 (7%)	0 (0%)	15 (100%)
Schools with PSAE and both semester 1 and semester 2 Biology exam	3 (100%)	0 (0%)	0 (0%)	0 (0%)	3 (100%)
Schools with PSAE and either semester 1 and semester 2 Biology exam	29 (43%)	21 (31%)	13 (19%)	2 (3%)	68 (100%)

Table 19: Performance Third Comparison of Schools with PSAE and Schools with PSAE and Biology Exam On File

CPS High Schools	School Size			Total
	Large	Medium	Small	
Schools with PSAE	13 (16%)	48 (60%)	19 (24%)	80 (100%)
Schools with PSAE and semester 1 Biology exam	10 (21%)	35 (73%)	3 (6%)	48 (100%)
Schools with PSAE and semester 2 Biology exam	2 (13%)	12 (80%)	1 (7%)	15 (100%)
Schools with PSAE and both semester 1 and semester 2 Biology exam	0 (0%)	3 (100%)	0 (0%)	3 (100%)
Schools with PSAE and either semester 1 and semester 2 Biology exam	20 (31%)	40 (62%)	5 (7%)	65 (100%)

Table 20: Size Comparison of Schools with PSAE and Schools with PSAE and Biology Exam On File

Chi-Squared Tests

As mentioned previously, applying a chi-squared test is an incomplete test to determine statistical independence, as the sampling method to collect the exams is not a simple random sample. Still, we can use this analysis to demonstrate that within typical statistical considerations, the high schools with biology exams are independent from all high schools. Let us first consider the data shown on Table 17.

I will first state the null hypothesis and an alternative hypothesis.

H_0: Biology exam presence and poverty levels are independent.

H_a: Biology exam presence and poverty levels are not independent.

I will next apply the chi-squared test for independence. Using Microsoft Excel's built in chi-squared test and *p*-value functions, the results shown on Table 21 are obtained. It is clear that the resulting *p*-values are greater than a significance level of 0.05, so the alternative hypothesis cannot be accepted. Thus, biology exam presence and poverty levels are independent.

On Table 22, *p*-value results for the exam distribution versus performance level are shown, again with the null hypothesis that exam presence and performance level are independent. On Table 23, *p*-value results for the exam distribution versus performance third are shown, again with the null hypothesis that exam presence and performance third are independent. On Table 24, *p*-value results for the exam distribution versus school size are shown, again with the null hypothesis that exam presence and school size are independent. None of these tests resulted in a *p*-value less than the significance level, so the alternative hypothesis cannot be accepted in any case. Therefore, in all cases, the presence of a Biology exam is independent from the school characteristic variable.

Null Hypothesis	**Chi-squared value**	***p*-value**	**Accept alternative hypothesis?**
Schools with PSAE and semester 1 Biology exam and poverty are independent	0.00147	0.5	No
Schools with PSAE and semester 2 Biology exam and poverty are independent	2.17×10^{-12}	0.5	No
Schools with PSAE and both semester 1 and semester 2 Biology exam and poverty are independent	8.009×10^{-17}	0.5	No
Schools with PSAE and either semester 1 and semester 2 Biology exam and poverty are independent	0.2066	0.4	No

Table 21: Results of Chi-Squared Tests On Exam Distribution Versus Poverty

Null Hypothesis	**Chi-squared value**	***p*-value**	**Accept alternative hypothesis?**
Schools with PSAE and semester 1 Biology exam and performance level are independent	0.0016	0.5	No
Schools with PSAE and semester 2 Biology exam and performance level are independent	9.48×10^{-13}	0.5	No
Schools with PSAE and both semester 1 and semester 2 Biology exam and performance level are independent	4.22×10^{-17}	0.5	No
Schools with PSAE and either semester 1 and semester 2 Biology exam and performance level are independent	0.0095	0.05	No

Table 22: Results of Chi-Squared Tests On Exam Distribution Versus Performance Level

Null Hypothesis	**Chi-squared value**	***p*-value**	**Accept alternative hypothesis?**
Schools with PSAE and semester 1 Biology exam and performance third are independent	0.0022	0.5	No
Schools with PSAE and semester 2 Biology exam and performance third are independent	2.81×10^{-12}	0.5	No
Schools with PSAE and both semester 1 and semester 2 Biology exam and performance third are independent	5.09×10^{-16}	0.5	No
Schools with PSAE and either semester 1 and semester 2 Biology exam and performance third are independent	5.92×10^{-16}	0.5	No

Table 23: Results of Chi-Squared Tests On Exam Distribution Versus Performance Third

Null Hypothesis	**Chi-squared value**	***p*-value**	**Accept alternative hypothesis?**
Schools with PSAE and semester 1 Biology exam and school size are independent	0.00014	0.5	No
Schools with PSAE and semester 2 Biology exam and school size are independent	2.59×10^{-12}	0.5	No
Schools with PSAE and both semester 1 and semester 2 Biology exam and school size are independent	7.77×10^{-17}	0.5	No
Schools with PSAE and either semester 1 and semester 2 Biology exam and school size are independent	0.00045	0.5	No

Table 24: Results of Chi-Squared Tests On Exam Distribution Versus School Size

Conclusion: This Sample Is Representative

In this section, I have provided evidence to support the following claims about the set of high school exams collected as part of this research. The schools that have biology exams on file are representative of the entire CPS high school population, so claims about this population are valid claims for CPS as a whole. Sub-population claims (school size, school poverty level, and school performance level) based on this sample population are also valid claims for CPS as a whole. The collection of semester 1 and semester 2 exams are representative when taken independently, but despite the chi-squared test results for schools with both semester 1 and semester 2, I will not conclude this content represents the complete set of content learned by students over the course of their two-semester biology course.

Focus Group

With these data tables and some preliminary analysis, I conducted a focus group of CPS science teachers to elicit their interpretation of the semester exam in CPS science classrooms.

Teachers were recruited via the CPS science email listserve. A copy of the advertisement to recruit teachers is shown on Figure 3. Anyone interested in participating was invited, resulting in a focus group that is not an elite group of teachers, but also one that likely is not particularly average, gifted, or representative of anything other than high school science teaching throughout CPS.

Recruitment Advertisements

The potential participants for focus groups will be recruited at random from CPS science teachers who participate on the CPS email mailing list. For more information on the CPS email mailing list, please visit http://cmsi.cps.k12.il.us/department_mailinglists.asp. (This email list has many suburban teachers who subscribe as well, and is frequently cross-linked with various suburban science teaching organizations.)

Attention science teachers:

I'm recruiting experienced high school science teachers for a research activity. Selected teachers will be asked to participate in an 1.5-hour focus group and discuss findings related to the content, rigor, and assessment literacy inherent in teacher-created final exams from high school science courses. I'm looking for a balance of teaching experience, school type, content specialization, to ensure we generate a well-rounded perspective. The focus group will occur at [location] on [date] at [time]. Upon completion, participants will receive a $25 Amazon.com honorarium for participation. If you are interested in participating, please contact Michael Lach at mlach@teachandlearn.org or 773/841-0149 and provide the science courses you've taught, the number of years you've been teaching, and the school name in which you currently work.

More details:

The title of this research is "High School Science Semester Tests In The Chicago Public Schools," protocol number 2009-0099. By participating in this research project, you'll help advance our knowledge about the depth and breadth of science instruction in Chicago high schools. The principal investigator is an Ed. D. student at the University of Illinois at Chicago, and this research constitutes his doctoral thesis. A copy of the research protocol and methodology is attached.

The principle investigator is Michael Lach of 1705 West George Street, Chicago, IL 60657, mlach@teachandlearn.org, 773/841-0149.

Figure 3: Advertisement Sent to CPS Email Lists To Recruit Focus Group Participants

Twelve teachers were selected to participate, and nine actually did. All were practicing, CPS high school teachers from a variety of schools. All participants completed an informed consent form (see Appendix 2, page 149) and received a $25 Amazon.com gift card for their efforts.

During the focus group, participants responded to and engaged in a discussion driven by a series of slides (see Appendix 3, page 155) that outlined the study and emerging data. These slides included a set of questions that were embedded into the discussion.

The focus group conversation was recorded and then transcribed by subcontractors of Logos Consulting LLC, a small but well regarded local evaluation firm. The transcript was reviewed for key comments and insights, and these will be used as part of the analysis.

Ethics

The ethical issues of this research project mostly revolve around the typical human subjects research provisions that are governed by the university's Institutional Review Board. This research is exempt from IRB review for the following reasons. (Citations are from the US Health and Human Services Human Subjects website).

First, this is research that is "conducted in established or commonly accepted educational settings, involving normal education practices," leading me to believe that exemption 45 CFR 46.101(b)(1) applies. Creating semester and final exams in high school science classes is certainly a typical and commonly accepted practice, and the high school science classroom is indeed a "normal education practice." The data was aggregated during the course of the research, and it was not part of any evaluation process. Indeed, individual teacher names are not included in my original data set of final exams—only the course name and school.

Second, this is research that involves "…the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens." While other data have been collected (the focus groups; thus the need for the exemption described above), the exams that have been collected will not enable individual teachers to be identified.

Of course, it is paramount that in the course of this research, the rights of the teachers whose exams I have study are protected and preserved. There is no way that teachers could be put at risk of criminal or civil liability, or be socially or emotionally damaged from the course of this research.

On March 25, 2009, I was granted a claim of exemption through March 20, 2012. On February 11, I re-submitted this request, and was granted a new claim of exemption through February, 2015.

Chapter 5: Findings

Introduction

In chapter 3 I described the conceptual framework for this study and the methodologies used to collect and process the data. For this study, the primary data source is a collection of teacher-written end-of-semester final exams from a variety of Chicago Public School high school science courses. At the conclusion of chapter 4, I confirmed that the sample of exams that I have collected is generally representative of the system as a whole.

In this chapter I will analyze that data. My conclusions will be presented as a series of findings, with each finding supported by a different take on the data. I will organize my findings into four major sections—findings about the exams and exam design, findings about the item type and format, findings about content and content distribution, and findings about cognitive demand.

There are many, many possible arrangements of this data that would likely result in interesting findings about science instruction in Chicago high schools. I have settled on the analyses that are likely the most foundational and the most relevant to the largest audience. In Chapter 6 I discuss some of the limitations this engenders, and provide options for further research should more data and resources become available.

<u>Exams and Exam Design</u>

The first set of questions will look at the exam table, as described in chapter 3, and attempt the make some conclusions about how exams are structured and designed. As part of the data collection and coding process, the number of items on each exam was recorded. There are two major findings associated with exam length.

Finding 1: The number of items according to teachers is often not the number of responses required of students.

As described in chapter 4, in the process of coding exams, it became clear that the question number on the actual exam did not always represent the question number of the student response desired. For instance, a fill-in-the-blank question that is one sentence with two blanks is actually two student responses—and the student could presumably generate none, some, or all of the correct responses. On some exams, teachers count the sentence; on others they count each possible response. As shown on Table 25, almost one third of exams across all content areas exhibited this disconnect.

Did the teacher item count match the actual item count?	**Course**			**Total**
	Biology	Chemistry	Physics	
No	27%	27%	45%	29%
Yes	73%	73%	55%	71%
Total	100%	100%	100%	100%

Table 25: Actual Item Count versus Teacher Item Count

In the course of the exam document review, I did not obtain much information about how the exams were scored and how this raw score was turned into a student grade. In the focus group, teachers commented that points are often assigned differently for particular items, but that with the rise of computer-graded tests, the focus has shifted to getting a raw number from the machine to easily enter that into a grade book with minimal weighting. For instance, one focus group teacher commented, "…of course everything's worth the same because of course you want the machine to tell you what their percentage is and that's what you put down, you know" and the group nodded in agreement. At best, the dis-connect between teacher item count and student item count means that most exams had a relatively complex scoring guide (one point for these questions, two points for those ques-tions, etc.). However, few exams included detailed scoring guide or weights on the paper given to students. At worst, it indicates sloppy and imprecise assessment development. Given the high number of exams that exhibited this characteristic, it points to a relatively unsophisticated scoring practice overall, and an incomplete communication of the practice of turning correct and incorrect responses into grades to students.

Finding 2: By subject, all exams exhibited similar characteristics of length.

Statistics that describe the number of items on this set of exams are shown on Table 26, and a distribution is shown on Figure 4. Aside from the tendency of physics exams to be somewhat shorter than the others, the number of item length is similar overall but exhibits considerable variation. As shown on Table 27, Table 28, and Table 29, the average number of biology items on an exam does not vary much by poverty level, performance level, or size of the school.

In the focus group, teachers commented that generally exams were designed to take place during one class period—roughly 45 minutes long. As one teacher mentioned, “you give [students] like a 10-question worksheet, it takes them like 3 days to get it done,” so shorter exams were preferred.

Course	**Number Of Items**			
	Average	**Maximum**	**Minimum**	**Standard Deviation**
Biology	51.7	100	15	19.7
Chemistry	49.9	100	25	20.8
Physics	38.3	81	13	19.3
Grand Total	50.0	100	13	20.0

Table 26: Descriptive Statistics on Number Of Items In Exams

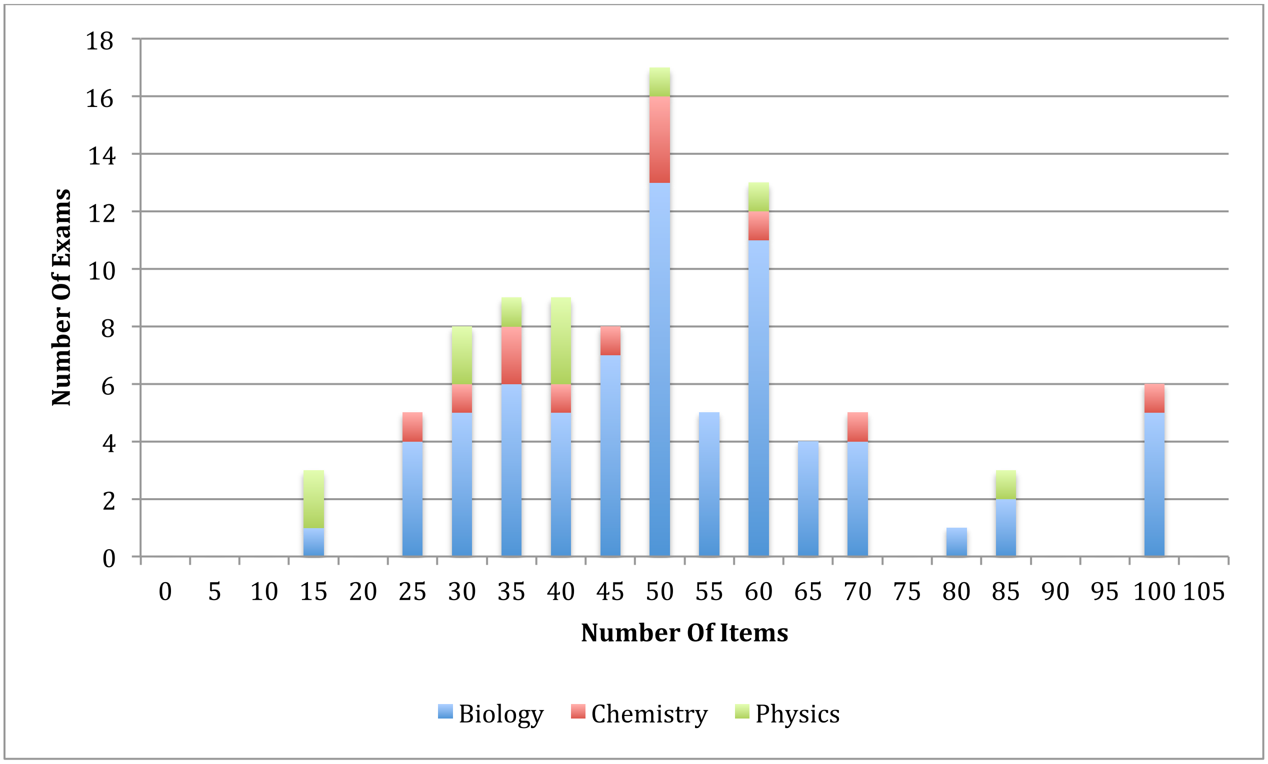


Figure 4: Frequency Count Of Number of Items

Poverty Level	Average Number Of Biology Items
High	52.2
Low	48.2

Table 27: Average Number Of Biology Items versus Poverty Level

Performance Level	Average Number Of Biology Items
Low	51.4
Middle	54.1
High	54.0
No Data	45.7

Table 28: Average Number of Biology Items versus Performance Level

School Size	Average Number of Biology Items
Large	51.8
Medium	51.3
Small	53.6

Table 29: Average Number of Biology Items versus School Size

Finding 3: Many schools have many exams for the same course.

As indicated in chapter 4, the 73 active biology exams come from 44 CPS schools. This means many schools had more than one exam per course, as shown on Table 30.

Number Of Biology Exams Per Schools	2003-04		2004-05	
	Semester 1	Semester 2	Semester 1	Semester 2
Schools with 1 exam	11	5	14	0
Schools with 2 exams	3	2	4	0
Schools with 3 exams	4	1	1	1

Table 30: Number Of Biology Exams Per Semester and Per School

The initial request made to schools to collect exams was for one exam per course. For nearly one third of the high schools, however, multiple exams for the same course were submitted. The implication here is that at these schools, there is not much curricular coherence between classrooms teaching the same course—some teachers assess different material in the same course than other teachers do.

The focus group confirmed how hard it is for teams of teachers at the school to reach consensus about what should be on their exam. One teacher commented, “I know in our school there’s lots of disagreement about what should be covered, so some teachers emphasize some topics more than others.” Another teacher elaborated further.

> *“Yeah, it was very contentious... to get groups of people to agree on one kind of test. The closest we got… was to say here’s the core principles that we're going to make for our first quarter test, and we're all going to give this core exam. And then other people said, well, I want to add more to it, so I covered this you guys didn’t cover, yeah, that's fine, but we’re all going to give the same core one.”*

One teacher highlighted the leadership issues at play.

> *“For example when I was the department chair we submitted one biology exam, but that's not to say that's the actual exam given by that person in the classroom. That was the official one. So someone in that case, they sent [CPS headquarters] the official one and [students] got the unofficial one.”*

Item Type and Format

I will now analyze the item type distribution on these exams. For this analysis, it is important to note (as described in the previous chapter and above) that item weights were requested on all exams but few were received. Where none were received, I assumed that each answer represented one point, but the actual (and important) process of converting item scores to grades is beyond the scope of this research.

Table 31 shows the proportion of item types across the entire set of items by course. By way of interpreting these numbers, the 1% in the Biology column and constructed response row means that of the entire active set of biology items, 1% of them were classified as constructed response. The grand total is the proportion across the entire data set irrespective of course.

Item Type	**Course**			**Grand Total**
	Biology	Chemistry	Physics	
constructed response	1%	0%	5%	1%
fill in the blank	3%	0%	2%	3%
matching	1%	0%	1%	1%
multiple choice	83%	91%	64%	82%
rewrite sentence	0%	0%	2%	0%
short answer	7%	8%	19%	8%
true or false	4%	1%	7%	4%
Grand Total	100%	100%	100%	100%

Table 31: Item Type Distribution By Content

When biology item types are examined across semester and school year, no real differences emerge, as shown on Table 32, with the exception of semester 2 exams in 2004-05, which show slightly more constructed response and short answer items.

Item Type	Biology Exam Items				Grand Total
	2003-04		2004-05		
	Semester 1	Semester 2	Semester 1	Semester 2	
constructed response	0%	3%	1%	6%	1%
fill in the blank	5%	2%	2%	0%	3%
matching	2%	0%	1%	0%	1%
multiple choice	83%	86%	83%	72%	83%
no match	0%	0%	0%	0%	0%
short answer	6%	8%	6%	22%	7%
true or false	5%	0%	6%	0%	4%
Grand Total	100%	100%	100%	100%	100%

Table 32: Biology Item Distribution By School Year and Semester

Finding 4: Most items are multiple-choice.

A comparison of these two tables shows the vast majority of items are multiple choice; constructed response and short answer questions tend to be worth more proportionally on exams; and exams tend to be similarly constructed across content areas.

The focus group conversation provides three reasons that these item distributions match what schools are looking for.

First, principals want exams that align to the ACT exam, the high-stakes exam for their schools. But alignment here most often means aligned by format more than anything else. “You do try to make [your exam] ‘ACT style,’” one teacher commented, because “principals care deeply about [standardized] assessments.” The rationale seems to be that students need to get used to the format and style because “the significant [exams] are multiple choice”—almost as if preparing students for exam format trumps exam content. There seems to be a shared belief among teachers and their administrators that high scores are as much a function of being familiar with the format as knowing the content. “You should have stuff that models” the ACT format was the consensus of teachers. Teachers feel the pressure here, too: “there’s so much pressure on ACT scores.”

The second push for multiple-choice came because of the time. Constructed response exams take time to grade. Multiple-choice exams—particularly with automated scoring systems—take much less time. As one focus group member commented, “…because now we have the Scantron exams, so we can do more stuff online. So it'd be much easier to [score] now because we have one biology exam online.” In another teacher’s words:

> *“So I’m not surprised by that. Because even now, we only have a few constructed response, but they’re very time-consuming to grade. When you have one-hundred-and-fifty kids times four, that’s six hundred questions I’ve got to grade in a couple of days. That’s just why people say, if I don’t have to do it, if it’s all multiple choice, you run it through the machine, with the item analysis.”*

The third reason for the preponderance of multiple-choice items was because they are easier for students. Teachers commented how their students have very low literacy skills. For instance:

> *"…Because we have students write lab reports when they do lab, but the quality of their writing is so poor that, like, they can't convey what they know a lot of times, and so we started, in addition to the lab report, giving them like a lab quiz, like a multiple choice one, and they scored significantly better on the quizzes than they ever did on the lab reports."*

Even when teachers put lots of attention on writing and literacy skills, the progress is slow.

> *"[We worked really hard] to get the kids to do some rewriting, and writing to learn. And then we did a lot of constructive response practice. So now we're at the point where most of the kids do the constructive response, but their scores are very bad, they're still very low."*

And there is lots of pressure to pass students on. The "actual reality is that like I'm under tremendous [pressure], I'm in a freshmen academy, and our whole mission is to pass more freshmen." Another focus group teacher commented:

> *"So if my kids don't do good on constructed response, it'd be foolish for me to put a lot of those on the test, it's just going to make their scores lower. So if I know, and then I'm not saying we do this overtly, but maybe subconsciously, so if I know my kids have a better chance on multiple choice, I'm going to give them more multiple choice, because then I'm going to be able to pass more kids. And I mean I hate to be that overt, but that's, I'm under a lot of pressure to make sure my failure rate doesn't go above or below a certain point. And so we're going to try to give them an assessment that they're going to do the best on."*

Finding 5: Physics exams have more short answer and constructed response items than biology or chemistry exams.

The one exception to this is that physics exams tend to have more short answer questions (often simple calculation or application of formulas) than biology or chemistry.

Focus group teachers thought this was reasonable, though they did not have really good reasons for it:

> *"Unfortunately, you know, from my experience, you know, chemistry and biology exams tend to be [lots of] multiple choice questions, and physics exams, at least in my experience, have tended to be more constructed response."*

Content Coverage And Distribution

Now, I will examine the content that is covered on these exams.

Focus group teachers confirmed that final exams are a good measure of what is actually taught in their classrooms.

> *"What I notice in my students is generally with my students, a C student will generally get a C on the exam so I feel like I exams are pretty reflective of at least their content knowledge and their skills knowledge that we're asking for even though there's lots of things that we do in the classroom."*

In the coding process, reviewers categorize the content on each item of each exam using the *Survey of Enacted Curriculum* content categories. (See chapter 4 for more information, as well as appendix 1 on page 144 for specific details on the *Survey of Enacted Curriculum* materials.) Each reviewer could choose up to three content categories for each item, and in cases where multiple reviewers examined the exam, only those categories indicated with consensus were kept. In this analysis, I only report on the high level "K-12 Science Content Areas" and not the finer-grain sub-codes in order to reduce complexity, but as noted in the final analysis, a sub-code analysis might reveal some different findings.

Finding 6: Most items only focus on one particular piece of content.

On Table 33, the number of unique content codes is shown by course. From this table, it is clear that most items assessed one particular piece of content. Given the preponderance of multiple-choice items on these exams, the fact that most items only assess one particular piece of content is not too surprising—crafting multiple-choice questions that address multiple aspects of science is difficult. Items that have zero unique content codes assessed content that in the coder's perspective was not part of the *Survey of Enacted Curriculum's* science coding framework.

Course	**Number of Unique Content Codes**				**Grand Total**
	0	1	2	3	
Biology	294	2803	587	92	3776
Chemistry	0	459	109	10	578
Physics	44	226	125	27	422
Grand Total	338	3488	821	129	4776

Table 33: Number of Unique Content Codes

Content Coverage By Course

Now, I will examine the content covered by particular courses in schools—looking at exams more holistically to determine the content they address. Since students experience whole exams as a collection of specific items, this is likely the most relevant way to explore the science content teachers expect their students to learn.

Making Sense Of Content Codes

Of the 4776 total items, 35 items did not get matched by coders and were rejected. Furthermore an additional 338 did not get scored for content as mentioned above (content that was not on the *Survey Of Enacted Curriculum* content guide, often questions about specific classroom or school procedures), so the total number of items under consideration here is 4776 - 35 - 338 = 4403.

As described in Chapter 4 and shown on Table 33, the coders had the option to assign multiple content codes to some items. This is understandable, as the distinctions between aspects of content are somewhat fuzzy and demand the knowledge of the coder be brought to bear on the decision. For biology, 2803 items had one code, 587 had two codes, and 92 items received three codes. For this analysis, I will treat each code as a separate item and display results as a proportion of the whole. Thus, for biology, while there are 3776 total items, there are 2803 + 2 × 587 + 3 × 92 = 4259 total codes. As mentioned in chapter 4, if the item weight was unknown I assumed equal weighting across items. The proportion of items with no match and no code are also listed here for completeness.

I expect different coverage in first semester and second semester, which gives us the proportional coverage by topic as shown in Table 34. To interpret this table, the 6% for the Nature of Science row and Biology Semester 1 column means that 6% of the assigned content codes in the entire data set for biology semester one exams were Nature of Science. Table 34 paints a picture of content across the entire data set.

Content Coverage Over Entire Data Set	**Biology**		**Chemistry**		**Physics**	
	S1	S2	S1	S2	S1	S2
Nature of Science	6%	3%	4%	0%	2%	0%
Science & Technology	2%	0%	2%	4%	0%	0%
Science, Health & Environment	1%	0%	0%	0%	0%	0%
Measurement & Calculation in Science	2%	1%	12%	6%	16%	5%
Components of Living Systems	28%	8%	0%	0%	0%	0%
Biochemistry	11%	4%	0%	0%	0%	0%
Botany	6%	1%	0%	0%	0%	0%
Animal Biology	1%	1%	0%	0%	0%	0%
Human Biology	1%	0%	0%	0%	0%	0%
Genetics	7%	17%	0%	0%	0%	0%
Evolution	6%	15%	0%	0%	0%	0%
Reproduction & Development	5%	3%	0%	0%	0%	0%
Ecology	8%	18%	1%	0%	0%	0%
Energy	0%	0%	3%	0%	14%	25%
Motion & Forces	0%	0%	0%	1%	57%	70%
Waves	0%	0%	0%	0%	2%	0%
Kinetics and Equilibrium	1%	0%	2%	19%	2%	0%
Properties of Matter	5%	3%	26%	16%	0%	0%
Earth Systems	0%	0%	0%	0%	0%	0%
Astronomy	0%	0%	0%	0%	0%	0%
Meteorology	0%	1%	2%	0%	0%	0%
Elements & The Periodic System	0%	0%	7%	1%	0%	0%

Content Coverage Over Entire Data Set	Biology		Chemistry		Physics	
	S1	S2	S1	S2	S1	S2
Chemical Formulas & Reactions	3%	1%	38%	26%	0%	0%
Acids, Bases, & Salts	2%	2%	1%	18%	0%	0%
Organic Chemistry	2%	0%	0%	8%	0%	0%
Nuclear Chemistry	0%	2%	2%	0%	0%	0%
no match	1%	3%	0%	0%	0%	0%
no score	4%	17%	0%	0%	8%	0%

Table 34: Content Coverage By Course

But this table (Table 34) describes the distribution of items across the whole sample. Also relevant is the proportion of exams that include this content—that provides an indicator of the distribution of content that students are expected to learn at particular schools system wide. Table 35 shows the proportion of exams that include at least one item addressing a particular content area. To interpret this, the 71% Nature of Science row in the Biology semester 1 column means that of all 58 biology semester 1 exams, 71% included items that in addressed Nature of Science.

Proportion Of Exams Featuring Particular Content	Biology		Chemistry		Physics	
	S1	S2	S1	S2	S1	S2
	58 exams	15 exams	7 exams	5 exams	10 exams	1 exam
Nature of Science	71%	27%	57%	27%	40%	0%
Science & Technology	41%	0%	29%	0%	0%	0%
Science, Health & Environment	10%	13%	0%	13%	0%	0%
Measurement & Calculation in Science	36%	7%	71%	7%	90%	100%
Components of Living Systems	86%	33%	0%	33%	0%	0%
Biochemistry	78%	40%	14%	40%	0%	0%
Botany	72%	20%	0%	20%	10%	0%
Animal Biology	16%	20%	0%	20%	0%	0%
Human Biology	29%	13%	0%	13%	0%	0%
Genetics	50%	73%	0%	73%	0%	0%
Evolution	62%	80%	0%	80%	0%	0%
Reproduction & Development	52%	40%	0%	40%	0%	0%
Ecology	53%	67%	14%	67%	0%	0%
Energy	10%	0%	43%	0%	60%	100%
Motion & Forces	2%	0%	0%	0%	90%	100%
Waves	3%	0%	0%	0%	10%	0%
Kinetics and Equilibrium	12%	7%	14%	7%	30%	0%
Properties of Matter	64%	13%	86%	13%	10%	0%

Proportion Of Exams Featuring Particular Content	Biology		Chemistry		Physics	
	S1	S2	S1	S2	S1	S2
	58 exams	15 exams	7 exams	5 exams	10 exams	1 exam
Earth Systems	9%	7%	0%	7%	0%	0%
Astronomy	3%	0%	0%	0%	0%	0%
Meteorology	3%	13%	14%	13%	0%	0%
Elements & The Periodic System	9%	0%	86%	0%	0%	0%
Chemical Formulas & Reactions	62%	20%	100%	20%	0%	0%
Acids, Bases, & Salts	52%	13%	43%	13%	0%	0%
Organic Chemistry	38%	13%	0%	13%	0%	0%
Nuclear Chemistry	9%	20%	29%	20%	0%	0%

Table 35: Proportion Of Exams With Questions In Particular Content Areas

Several findings emerge from these tables.

Finding 7: There is lots of content diversity among biology exams, and less in chemistry or physics.

The diversity of topics that are included in biology exams is quite striking. Only four topics—motion and forces, waves, astronomy, and meteorology—appear on less than 5% of the first semester biology exams. Compare this to first semester chemistry, which has 13 topics that do not appear, and first semester physics, which has 18.

Finding 8: Biology content follows a traditional sequence, with science process concentrated in the first semester.

While there is lots of diversity in biology exams, course content—as measured by the most frequently reported content on exams—still falls into traditional course content descriptions. On Table 36, Table 37, and Table 38, the most frequently reported topics by semester are reported for each course.

Semester 1	Semester 2
Components of Living Systems Biochemistry Botany Nature of Science Properties of Matter Evolution Chemical Formulas & Reactions	Evolution Genetics Ecology

Table 36: Most Frequently Reported Biology Content

Semester 1	Semester 2
Chemical Formulas & Reactions Properties of Matter Elements & The Periodic System Measurement & Calculation in Science Nature of Science	Chemical Formulas & Reactions Acids, Bases, & Salts Science & Technology Kinetics and Equilibrium Properties of Matter

Table 37: Most Frequently Reported Chemistry Content

Semester 1	Semester 2
Measurement & Calculation in Science Motion & Forces Energy	Measurement & Calculation in Science Motion & Forces Energy

Table 38: Most Frequently Reported Physics Content

To understand this, I will compare the content contained in these exams with other references for standard high school Biology content.

In Biology, the sequence of topics traditionally utilized in classrooms came from a series of mutually reinforcing documents for teachers: national associations (National Science Teachers Association, 1996), from the district (Chicago Public Schools, 2000c), or from textbook publishers. The recommended course content for Biology in the Chicago Public Schools in 2000 was described by a document called *Expecting More* (Chicago Public Schools, 2000a) and summarized on Table 39. The course order as reported here was based on popular textbooks from Prentice Hall (Miller & Levine, 1998), Glencoe/McGraw Hill (Biggs, Kaskel, Lundgren, & Mathieu, 1998), and Scott Foresman (Strauss & Lioswski, 1998). It is important to note that the *Survey Of Enacted Curriculum* coding process uses particular terminology to help assign exam items to content areas, but these have clear parallels with the categories described in *Expecting More* as shown on Table 40.

These textbooks typically present something called "The Scientific Method" in the introductory chapters, devoid from content (N. G. Lederman, 1998). While textbooks that provide an alternative approach to this do exist (such as BSCS's *Biology: A Human Approach,* published by Kendall-Hunt) they are not very popular in schools and classrooms—

and likely will not be, as long as district documents promote an alternative and more popular arrangement.

Biology Content from *Expecting More*	
Semester 1	Semester 2
Chemistry of Biological Processes Cellular Structures and Functions Energy Processes in Living Things DNA and Protein Synthesis	Principles of Heredity and Reproduction Natural Selection and Evolutionary Adaptations Fundamentals of Taxonomy Environmental Cycles and Ecosystems

Table 39: Biology Content in CPS's *Expecting More* guide

When compared to the biology content reflected on exams as show on Table 35 and summarized on Table 36, clear parallels can be seen.

***Survey Of Enacted Curriculum* category**	**Likely Expecting More category**
Components of Living Systems	Cellular Structures and Functions Energy Processes in Living Things (cellular respiration and photosynthesis)
Evolution	Principles of Heredity and Reproduction Natural Selection and Evolutionary Adaptations
Ecology	Environmental Cycles and Ecosystems
Genetics	DNA and Protein Synthesis Principles of Heredity and Reproduction
Properties of Matter	Chemistry of Biological Processes

***Survey Of Enacted Curriculum* category**	**Likely Expecting More category**
Nature of Science	No match
Botany	Fundamentals of Taxonomy (plants)
Biochemistry	Chemistry of Biological Processes (likely not organic chemistry)

Ta**ble 40: Alignment between items on the Survey Of Enacted Curriculum and Expecting More**

The overall sequence of CPS exam content is quite similar to *Expecting More*—with a bit more chemistry, and a bit less taxonomy. The pacing seems slower than the recommended *Expecting More* pacing, which should not be too surprising—there are always complications at the school level that make pacing a challenge. Evolution is positioned as a first semester topic by CPS, but is one of three popular semester 2 topics on the exams.

Teachers in the focus group confirmed this observation. The teachers commented that most teachers do things in the same order: "people are going in the same order teaching that first semester, we all do it that way." We spend a lot of time on "the cell stuff, parts of cells," so it makes sense to see that so prominently reported. And "it also seems that people are going in the same order teaching that first semester. You know when you see genetics and evolution and ecology second semester, so obviously there's a lot of similarities among the schools."

As to the second semester topics, one teacher commented "I'm an evolutionary biologist. I know very few people get to that; it's the last chapter of most biology books so I'm really surprised it's on there at all." Others commented that this was a natural case of the realities of teaching hitting throughout the school year:

> *"At the beginning of the school year, you have all these wonderful goals and ideas and you set the tone and just do this classroom management, and it's hard to set the lab and your days quickly falls you know, it's easy to fall behind, you got some assemblies coming in there. It doesn't surprise me that there's more spread in a semester and that things slide."*

Another issue was the dependence of some content on previous content, which is probably more acute in chemistry and physics than in biology: "I also think because I just started teaching chemistry, and seeing that you can't do second semester chemistry without knowing first semester chemistry whereas like bio[logy], if you don't know the cell, you might be able to do ecology."

There was also the sense that we should expect this, given that teachers organize their classrooms the way they were taught and that change is difficult. One teacher noted "People tend to teach the way that they were taught. This very much looks like what I learned 20 years ago, sadly."

Finding 9: Science process skills are not assessed much.

Science process skills—even traditionally defined, like the scientific method, inquiry, the nature of science—seem to be predominantly taught in the first semester. While this is not very aligned to the Illinois Learning Standards and quite different from what is recommended in the *Next Generation Science Standards*, it does match the sequence common in basal textbooks (Strauss & Lioswski, 1998). It is perhaps not very surprising that this is presented this way—in contradiction to the modern approach of inquiry (Dow, 2000), the Illinois Learning Standards for science (Illinois State Board Of Education, 1997) put scientific processes into a separate standards (Illinois Learning Goal 11A) and do not connect them to content as current research recommends (Singer, Hilton, & Schweingruber, 2005).

Remarkably, the focus group teachers felt surprised to see this much science process at all—it was higher than they had expected. One teacher commented “the nature of science including the process of inquiry, understanding of the inquiry, [the fact] there is [a] significant amount of nature of science on [the exams], it’s exciting, I mean, I’d like to see more but, that’s pretty cool that they are, you know, asking, that they’re asking a lot of how to do science, not just on science.”

Finding 10: Some expected content is generally missing from Biology exams.

In particular the low frequency of human biology and animal biology topics is rather remarkable. Focus group teachers weren’t particularly surprised: “there’s so much to cover” was a popular refrain to questions about content choices.

Finding 11: There is little difference in content coverage between high poverty and low poverty high schools.

Now I will examine how content varies based on the poverty level of the school. Table 41 shows biology content coverage for schools by poverty level. This table includes exams from both semester one and semester two combined. A few things stand out from this table. First, the content coverage is remarkably similar—the frequencies that appear in each column are generally within a few percentage points of one another. Second, the low poverty schools tend to do less nature of science content—they appear to focus significantly more on traditional content than more progressive, ways-of-knowing type skills. Third, there is less introductory chemistry in the low poverty schools, perhaps because students there likely have a firmer foundation in the physical sciences upon entering.

Content Coverage	**Biology**	
	High Poverty	Low Poverty
	64 exams	9 exams
Nature of Science	67%	22%
Science & Technology	34%	22%
Science, Health & Environment	13%	0%
Measurement & Calculation in Science	30%	33%
Components of Living Systems	80%	44%
Biochemistry	72%	56%
Botany	66%	33%
Animal Biology	16%	22%
Human Biology	28%	11%
Genetics	53%	67%
Evolution	64%	78%
Reproduction & Development	52%	33%
Ecology	52%	89%
Energy	9%	0%
Motion & Forces	2%	0%

Content Coverage	Biology	
	High Poverty	Low Poverty
Waves	2%	11%
Kinetics and Equilibrium	13%	0%
Properties of Matter	56%	33%
Earth Systems	8%	11%
Astronomy	2%	11%
Meteorology	6%	0%
Elements & The Periodic System	8%	0%
Chemical Formulas & Reactions	56%	33%
Acids, Bases, & Salts	47%	22%
Organic Chemistry	36%	11%
Nuclear Chemistry	13%	0%

Table 41: Content Coverage by School Poverty Level

Teacher interpretations of this were mixed. Mostly, they agreed that the content should be similar—biology is biology, after all. They did note that in high poverty schools, "[students are] coming in with less of a science background, and it's so varied the back-ground that they're coming with. We have to start from ground zero and go over all of the basics."

__Finding 12: Both chemistry and physics appear to follow a traditional sequence as well.__

On Table 37 the most frequently categorized chemistry content is listed. On Table 38, the most frequently categorized physics content is listed. While the sample size here is quite smaller than that of biology, some interesting patterns are visible as well.

When compared to the CPS content guidelines for chemistry (Chicago Public Schools, 2000b) and physics (Chicago Public Schools, 2000c) it is clear that the sequence follows a similar flow to these guiding documents, which is not surprising.

In the focus group, teachers confirmed this sentiment. In physics, "you do kinematics; you do dynamics. You do Newton's laws. You do conservation of energy and the year is over." There just was not time to do anything else. The mathematics background of students can be a problem, too: "[Teachers] think math has to be taught because [students] don't know the skills, you know, they start teaching multiplication, all this you know, signing coefficients, etc."

__Finding 13: Some expected content is generally missing from Chemistry and physics exams.__

In particular the low frequency of organic chemistry, nuclear chemistry, and waves is remarkable. It is likely that most CPS students do not have much opportunity to learn this content. One focus group teacher commented, "I'm surprised that there's any high school doing organic chemistry, but then I'm surprised there's no biochemistry." The two physics teachers in the focus group both indicated they taught waves, but I suspect they were particularly high performing teachers.

Cognitive Demand

Now, I will explore how cognitive demand was distributed over these exams.

Making Sense Of Cognitive Demand Codes

Coders only assigned one cognitive demand code for each item. Again assuming the weight of items is equal on any given exam, this makes determining the relative proportion of cognitive demand items relatively easy. Of the total 4776 items, 22 of them had conflicting matches from coders, and so those items will be removed from the analysis. Furthermore, 21 items did not get scored, so those will also be removed, leaving a total of 4733 items to analyze.

When considering students and schools, the key consideration about cognitive demand is the degree to which levels of demand are distributed across an exam. So I will look at the proportion of items that fall into each of the cognitive demand categories and see how they compare across school type and discipline.

On Table 42, the proportions of items across the entire data set that fit a particular cognitive demand level per item per subject are shown. To interpret, the 74% at the intersection of the memorize row and the biology column means that 74% of the items that were on the 73 biology exams were coded at the memorize level of cognitive demand.

Cognitive Demand	Biology	Chemistry	Physics
	73 exams	12 exams	11 exams
Memorize	74%	36%	37%
Perform Procedures	13%	52%	45%
Communicate Understanding	6%	1%	6%
Analyze Information	6%	10%	9%
Apply Concepts/Make Connections	1%	0%	2%

Table 42: Average Proportion of Items With Cognitive Demand By Subject

Finding 14: Across all subjects, the proportion of exams with high levels of higher cognitive demand items is low.

As can be shown, the proportion of items with higher levels of cognitive demand is pretty low. 87% of biology exams focus on memorization and performing procedures; 88% of chemistry exams and 82% of physics exams focus in the same area. Very few items focus on applying concepts/make connections.

Focus group teachers attributed this to a variety of factors, most frequently the push towards multiple-choice questions because of grading time, a naïve understanding of "ACT style," and reliability. Focus group teachers were divided on the issue if multiple-choice items could even measure high cognitive demand—and they certainly agreed the writing items that could do that would be challenging.

Teachers in the focus group realized the contradictions they were espousing: "the high poverty kids need... analytical skills, which are most needed for constructive response pieces," but then countering with complaints about the low level of skills their students have and the lack of time they have to teach in this fashion. The consensus of the group was that the distribution of cognitive demand matched their experiences, but it was pretty frustrating for the focus group teachers to see.

Others said there was lots of pressure to pass students forward, particularly at the freshman level.

> *"We want to test what we're supposed to be teaching and what the kids actually know, but then the actual reality is that like I'm under tremendous [pressure]. I'm in a freshmen academy, and our whole mission is to pass more freshmen."*

There were some outliers, however. By taking a proportion of the number of weighted items with cognitive demand in categories D ("communicate understanding"), E ("analyze information"), and F ("apply concepts/make connections") over the total number of weighted items, we can see that there are a handful of exams that appear to have considerably more cognitive demand than others. A histogram of this information is shown on Figure 5. In Appendix 4 on page 170, a copy of an exam with a relatively high proportion of high cognitive demand items is shown (50%). In Appendix 5 on page 180, a copy of an exam with a low proportion of high cognitive demand items is shown. A review of these samples shows that there isn't much connection between items and item sets, despite the limitations of the *Survey of Enacted Curriculum* tool to identify such connections.

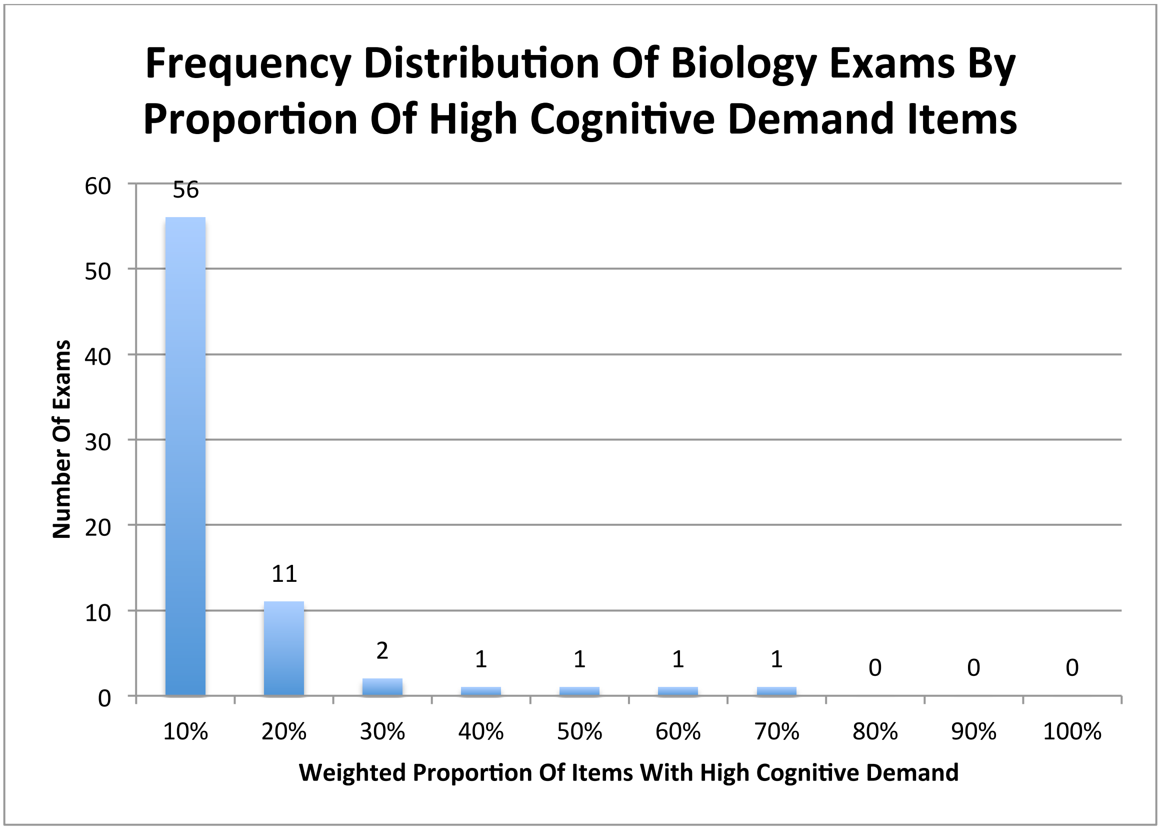


Figure 5: Frequency Distribution Of Biology Exams By Proportion Of High Cognitive Demand Items

Finding 15: Biology exams focus more on memorization than chemistry or physics.

Also from Table 42 and Figure 5, it is clear that there is a greater proportion of memorize items on Biology exams than on Chemistry or Physics exams. This makes sense given the typical sorts of questions that were asked on each. On biology exams, most of the items tended towards recall of a term or label of a process. In both chemistry and physics exams, most of the items were simple application of formulas (*e.g.* Newton's second law) to perform basic algebraic procedures.

The focus group teachers explained this in terms of vocabulary. One teacher mentioned that:

> *"One of the things that jumps out immediately is just the different between these [disciplines]. Biology for example is extremely vocabulary heavy, whereas chemistry you use like the same 20 vocabulary words the entire year so there's not that much to memorize."*

Another teacher echoed this: "There you have vocabulary in biology, a lot, and in physics as well, like 5 vocabulary words."

Yet another teacher attributed this to the fact that biology is predominantly a freshman level course, inadvertently highlighting a misconception about when and how children learn higher order thinking skills. "I think in most schools Biology is probably a freshman subject so that's why there might not be so many higher order thinking skills at that level, 'cause you're doing it with ninth graders."

Finding 16: Only slight differences among cognitive demand between high poverty and low poverty schools.

On Table 43, we can see the difference in proportion of items between high poverty and low poverty schools. In this case, there is a slight shift to more items at the "perform procedures" level, but it is not very different.

Cognitive Demand On Biology Exams	Low Poverty	High Poverty
Memorize	61%	80%
Perform Procedures	21%	10%
Communicate Understanding	6%	4%
Analyze Information	12%	5%
Apply Concepts/Make Connections	0%	1%

Table 43: Cognitive Demand Of Biology Items, By Poverty Level Of School

One of the focus group teachers had recently made a transition from a high-poverty school to a low-poverty school, and commented "I remember being at [school name], we were on probation so we had to do all this [test prep] stuff." Now, he has much more flexibility to assign more complex tasks and work to his students.

Chapter Conclusions

In this chapter, I have analyzed data about final exam structure, content, and cognitive demand to develop a picture of science assessment in Chicago high schools. To review, the major findings have been summarized on Table 44.

#	**Finding**
1	The number of items according to teachers is often not the number of responses required of students.
2	By subject, all exams exhibited similar characteristics of length.
3	Many schools have many exams for the same course.
4	Most items are multiple choice.
5	Physics exams have more short answer and constructed response items than biology or chemistry exams.
6	Most items only focus on one particular piece of content.
7	There is lots of content diversity among biology exams, and less in chemistry or physics.
8	Biology content follows a traditional sequence, with science process concentrated in the first semester.
9	Science process skills are not assessed much.
10	Some expected content is generally missing from Biology exams.
11	There is little difference in content coverage between high poverty and low poverty high schools.
12	Both chemistry and physics appear to follow a traditional sequence as well.
13	Some expected content is generally missing from Chemistry and physics exams.
14	Across all subjects, the proportion of exams with high levels of higher cognitive demand items is low.
15	Biology exams focus more on memorization than chemistry or physics.
16	Only slight differences among cognitive demand between high poverty and low poverty schools.

Table 44: Summary Of Findings

Taken together, these findings paint a particular picture of final exams across many Chicago science classrooms. The tests are long, and mostly low-level multiple-choice questions, because teachers feel pressure to defend their scores, want to pass students on, do not have time to grade more complicated items, and believe that matching formats is a path to ACT performance. The content follows traditional sequences, but does not completely cover what is expected. The differences among classes are slight, and likely due to understandings of courses that are grounded in tradition. Differences between high poverty and low poverty schools are low.

The content that teachers address aligns well with popular textbooks and the district guidelines, though teachers do not cover as much content as they would like to. The sequence and scope tends to follow that of popular textbooks, more so than either the Illinois Learning Standards or the *Next Generation Science Standards*. Despite claims about the importance of inquiry or the nature of science, those topics do not appear very much on exams. Evolution is included while the environment is not, but this seems to be more a function of sequence than any decision to address or avoid controversy. Teachers make decisions about content based on their own understanding of the disciplines and the resources they have available to them.

The "academic press" as explained by Lee *et. al.* (Lee, Smith, Perry, & Smilie, 1999) embodied with these exams is quite low. Most questions are recall or simple procedures, and there is little higher-level thinking or communication required. (A few outliers do exist, however.) This seems sadly intentional by teachers—a result of both unsophisticated understandings of the discipline, a discouraging sentiment that students are not capable of much more, and a workload that would make grading more complex exams overwhelming.

These are not sophisticated exams. Overall, there are few science educators who would want their own children in classrooms bookended by exams such as these. Yet they appear well designed for the context they operate in: teachers with limited content knowledge, classroom tools and central office guidance that are unsophisticated, and leadership that has a challenge bringing teachers together within the same school. While it is clear that this represents a pretty low level of assessment literacy on nearly any measure, it is also clear that the decisions that teachers are making are very much based on their experiences and contexts.

In the next chapter, I will discuss the probable causes of this situation and the implications for practice and research.

Chapter 6: Interpretations

Introduction and Overview

Chapter 5 concluded with a set of findings (reprinted from Chapter 5 as Table 44, below) based on an analysis of a large number of high school science end-of-semester exams from one large urban school district, the Chicago Public Schools. A discussion there maps those findings back to the research questions that drive this investigation.

In this chapter, I will identify the most important meanings or lessons from these findings by comparing them to broader issues in science education and education reform. I will connect them to some national data to explore the degree to which the Chicago Public Schools setting is unique, and attempt to distill that into the thinking that may well pervade most CPS science teachers. I will discuss why we see these findings, and the implications for the system as a whole, as well as for the Next Generation Science Standards. I will then conclude with some implications for practice that identify actions that teachers and leaders at all levels might utilize, and discuss the implications for further research. This discussion will include some literature review, some personal and policy reflection, and some evidence-based speculation.

#	**Finding**
1	The number of items according to teachers is often not the number of responses required of students.
2	By subject, all exams exhibited similar characteristics of length.
3	Many schools have many exams for the same course.
4	Most items are multiple-choice.
5	Physics exams have more short answer and constructed response items than biology or chemistry exams.
6	Most items only focus on one particular piece of content.
7	Biology exams demonstrate considerable content diversity. There is considerable, but less, content diversity in chemistry or physics.
8	Biology content follows a traditional sequence, with science process concentrated in the first semester.
9	Science process skills are not assessed much.
10	Some expected content is generally missing from Biology exams.
11	There is little difference in content coverage between high poverty and low poverty high schools.
12	Both chemistry and physics appear to follow a traditional sequence as well.
13	Some expected content is generally missing from Chemistry and physics exams.
14	Across all subjects, the proportion of exams with high levels of higher cognitive demand items is low.
15	Biology exams focus more on memorization than chemistry or physics.
16	Only slight differences among cognitive demand between high poverty and low poverty schools.

Table 45: Summary Of Findings

The Most Important Messages From These Findings

So what are the most important messages from these findings, particularly since they are based on data that is nearly 10 years old? As mentioned in chapter 3, despite the age of the data, this study provides an important benchmark that documents science teaching and learning in Chicago high schools at a particular point in time—as well as highlighting a method of obtaining such information in the future—that while limited, is cost effective and conducive to comparison studies across time and other districts.

First, **science teachers focus on a specific core of content** in their classes that is common across the system. There is a clear and consistent scope, sequence, and depth of content in each of the major science courses. While some exams focused a bit more on some topics than others, by and large biology exams across the city represented very similar content. Chemistry and physics exams demonstrated a similar core of content. This consistency occurred despite school type and population type. This core of content is quite different from the practices, crosscutting concepts, and disciplinary core ideas that make up the Next Generation Science Standards (NGSS Lead States, 2013b). In particular, science process skills got relatively little attention on teacher-generated exams.

Second, this **content is driven by both district guidelines and popular textbooks**, though the latter likely has more influence and both come from a long tradition of typical content. Both have a rather unsophisticated view of science teaching and learning, and a particularly simplistic view of scientific inquiry and processes. Given that high school science support in the Chicago Public Schools was minimal during the timeframe these end-of-course exams come from and the timeframe immediately prior (Wenzel, 2010), it is safe to assume that this was a function of both sources influencing practice: while teachers likely viewed the district documents as an authority, it was not a very far reaching one, and their local textbooks were the most immediate and concrete source of feedback. Standards seem to play only an indirect or incomplete role here, as issues like inquiry that receive extensive attention in the standards document get relatively little attention from teachers. Low expectations for students and a lack of push from school leadership enhanced the utility of these weak supports, creating a vicious cycle of instruction that concentrates on memorization and coverage. Later attempts to change instruction in Chicago Public School high school science courses via curriculum change highlighted how deeply these behaviors are ingrained into the culture of high school science teaching (Sporte et al., 2009).

Third, to most teachers, **rigor means coverage instead of intellectual press**. There are numerous content areas addressed on these exams, but they do so with generally low levels of intellectual demand. The adage that content in American classrooms is “a mile wide and an inch deep” matches that data found here. The dearth of high cognitive demand items is disappointing, as science educators have been talking about “critical thinking” and “higher level thinking” for decades (DeBoer, 1991). But in the Chicago high school science case, it is clear that rigor means more memorization and facts—exams could be quite long, and focused more on minute details of scientific trivia than general sense-making. The preponderance of low-level questions requiring factual recall or simple algebraic manipulation makes the occasional protestations from focus group teachers that other classroom activities—such as classroom laboratories and discussions—focus on more rigorous work rather suspect. And while the number of items that address inquiry or nature of science issues is quite low, to the focus group teachers, the proportions were surprisingly high. As one focus group teacher commented, "[the fact] there is [a] significant amount of nature of science on [the exams], it's exciting, I mean, I'd like to see more but, that's pretty cool that they are... asking a lot of how to do science, not just on science." In general, science teachers in the Chicago Public Schools are not asking their students to think very much on their exams.

Fourth, the **assessment seems to be delivering what is expected of it**. In the focus group, teachers indicated that final exams are a good measure of what is actually taught in their classrooms, but also expressed considerable frustration at the entire final exam routine. Teachers are pushed to use "ACT style" questions, which are interpreted as multiple-choice options. There is a deep teacher mistrust of the high-stakes accountability exams, as if students just need to be familiar with particular question types instead of deeply knowledgeable of content in order to score well (E. Allensworth, Correa, & Ponisciak, 2008a). Moreover, this occurs in a professional setting where they are not particularly trusted to use their own judgments about students and their learning. The teaching load of most classroom teachers (often more than one hundred students per teacher per semester) makes more time-intensive prompts and tasks logistically impossible or at least highly unlikely. In this context, a machine-scored collection of low-level multiple-choice questions seems disturbingly appropriate. Essentially, final exams are not enacted as they are designed. Teachers agreed that the purposes of final exams were generally laudable—provide an end-of-semester summative evaluation of overall student learning in the course. But the constraints placed on the routine, including a push for multiple-choice items, lots of pressure to pass students, little direction about what the exams should constitute, and little time for students to take or teachers to grade the exams, made it nearly impossible for teachers to craft final exams that did as intended.

Fifth, **the district is homogenous.** Across the major descriptors of schools—poverty level, size, and performance—biology exams were remarkably similar. The smaller collections of physics and chemistry exams were similar to one another as well. There were not pockets of schools with dramatically different exam structures and content. While there were a few exams that had items with more cognitive demand than others, 90% of the biology exams reviewed had 20% or less of their weighted items focused on higher cognitive demand, and 77% of the exams had less than 10%. As mentioned in chapter 5, this analysis focused on the higher-level "K-12 Science Content Areas" (at the hundreds level on the *Survey of Enacted Curriculum*) rather than the more fine-grained sub-codes (at the ones level). A more fine-grained analysis might have discerned more notable content difference between exams.

How should we reconcile this observed homogeneity of exams with the observed heterogeneity of performance? First, it's important to recall that science performance in Chicago high schools is not very heterogeneous. As shown on Figure 6, the majority of CPS high schools have 30% or less of their students meeting or exceeding standards on the Prairie State Achievement Exam (PSAE) in science in 2005, and only 14% of schools have 50% or more students meeting or exceeding standards. As shown on Figure 7, this isn't due to wide ranges of student performance within schools, as the proportion of students exceeding standards in most schools is also low. Aside from a few schools, most performance in Chicago Public Schools high schools was homogeneously low. And while the level of performance many be due to factors other than poor instruction, such as poverty or poor preparation, there is compelling evidence from other sources that the overall level of instruction in CPS high schools needs to improve (Chicago Public Schools, 2002).

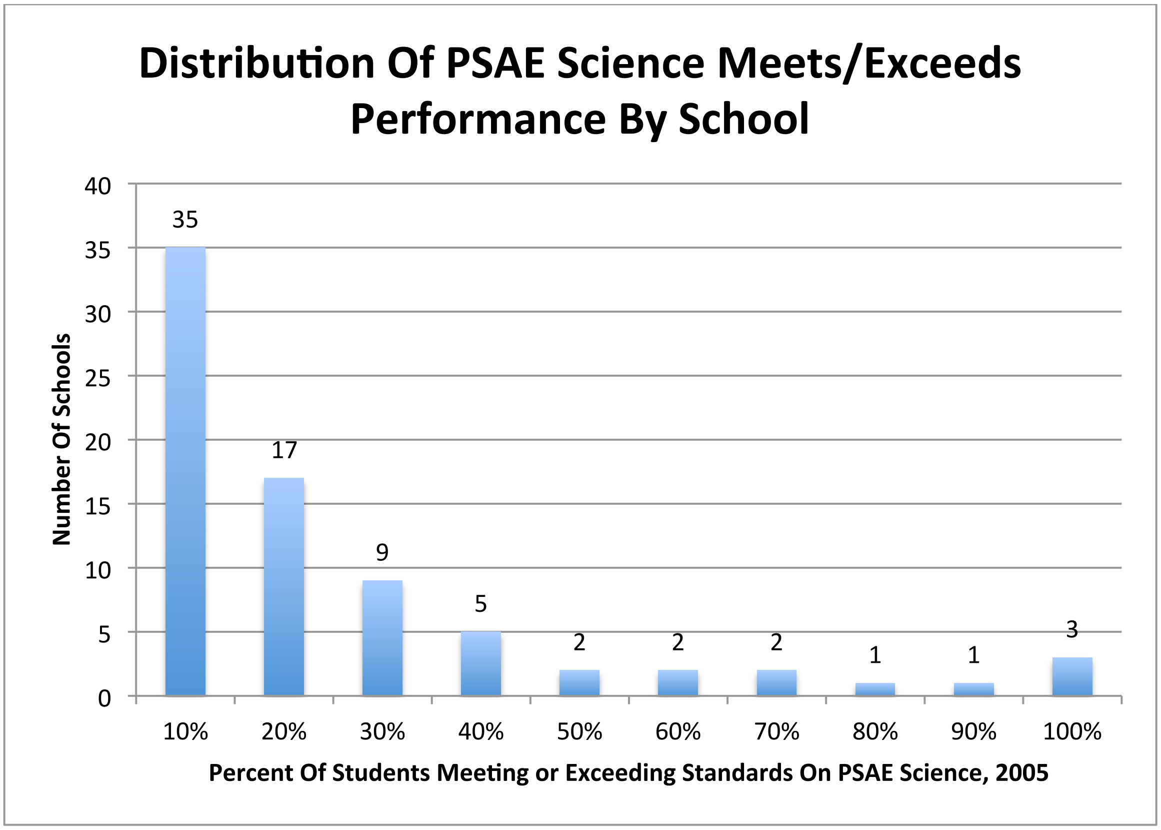


Figure 6: Frequency Distribution of PSAE Science Meeting/Exceeding Standards Performance By School

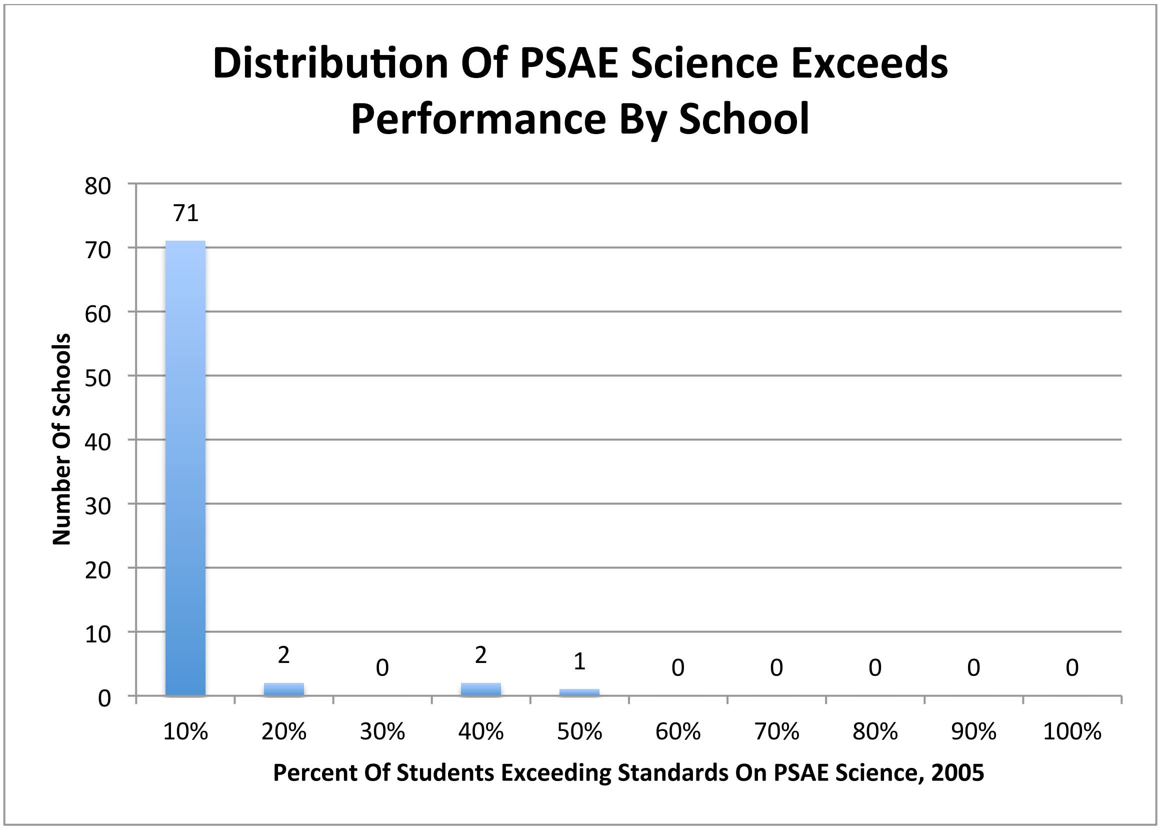


Figure 7: Distribution of PSAE Science Exceeds Performance By School

Secondly, while the PSAE is receptive to robust instruction (E. Allensworth et al., 2008a), the various components of PSAE science exam may not be equally so—and likely depend on reading ability a great deal (Ponisciak, 2005). Furthermore, given that biology is typically a grade 9 course (LaForce & Feranchak, 2003) and the PSAE is given in grade 11, it's unclear how sensitive PSAE scores are to biology instruction in particular. Unfortunately, a district-wide biology course exam that could be used to compare final exam quality with performance does not exist.

The implication of these last two messages is that there is a district culture among science teachers and within the final exam routine that drives these low levels of content coverage and rigor. Given the size of the district, the diversity of students, the diversity of school size, and the intention of final exams to provide a final summative evaluation of a semester of coursework, the lack of variance in final exams was unexpected. Given the limited evidence of the district's efforts to change this culture in any substantial way—by more supports, clearer policies, professional development, and curriculum materials—it is likely this culture runs deep. It is also likely that a combination of preparation efforts and local school culture contribute to building and sustaining this culture.

Despite these two key messages, we cannot clearly discern causes for them from this data:

On one hand, given that the final exam routine as enacted doesn't match how it is intended to function, teachers may create exams that are quite different from their overall instruction and the design of the course. Following this argument, teachers realize the futility of doing what they should do to with their finals, and the result is easy-to-write, easy-to-grade, easy-to-take tests for students—just as the date report here shows. They take the charge seriously—exams are a "big deal," after all, but can't reconcile it with the demands placed upon them and so compromises are made. Since assessment is always a practice of "design within constraints," in this context the constraints result in a broad content, low rigor exam—despite ostensible claims to students that it's truly summative.

On the other hand, the broad content coverage and low levels of rigor observed could point to significant lack of instructional capacity—that teachers don't have the tools, support, or know-how to design and administer tests that better represent the science content they are expected to teach. Because of the observed homogeneity of exams, it is likely not particularly a poverty problem, or a resource problem—though improving the situation would likely require additional resources. It is a problem with the culture that exists within the schools and the system. And given that other research shows that disadvantaged students tend to receive less effective teaching (Max & Glazerman, 2014), the implications here are sobering: There just are not the opportunities to learn science for students in the Chicago Public Schools that we might reasonably expect or that our city needs. Of course, both of these could be true as well—the Chicago Public Schools suffers from both an ill-designed final exam routine as well as low levels of instructional capacity—and the solution is to tackle both.

Connecting These Findings To The PSAE

As mentioned in chapter 2, the PSAE exam at the time these exams were collected consisted of two parts—a "day one" that was the ACT Science Reasoning exam and a "day two" that was an Illinois State Board of Education developed assessment of science content (Illinois State Board Of Education, 2006). The content focus on the collected exams is not dissimilar to the construction of the day two test: content-focused questions that generally concern themselves with science facts and minutiae. (The ACT Science Reasoning portion consisted of considerably more high cognitive demand items.) It is conceivable that this focus was driving teachers to make their exams similar. However, a cursory analysis of exams from courses that generally would be offered outside the typical course sequence, such as Astronomy, shows similar findings. The Astronomy exams generally consisted of recall-focused questions about astronomical terms and phenomena, in a multiple-choice format.

More research would be needed to determine definitely that the pattern holds, but an initial examination suggests that it does. Since these exams would generally be administered in senior year, after students had taken the PSAE, it indicates that it is likely that other factors were more important to teachers in designing their exams than the PSAE "day two" format.

Putting These Findings In A National Context

How does this differ from reports around the nation? There are not many similar studies with which to compare this data. Certainly, in terms of tested outcomes, science performance in the Chicago Public Schools is not so different from that in other large urban districts—though the most reliable comparison of science performance only addresses grade 4 and grade 8 (National Center for Education Statistics, 2011). The 2012 Survey of Science and Mathematics Education, conducted periodically by Horizon Research, Inc., is likely the most immediate and relevant comparison (Lyons, 2013). This work describes the status of high school (grades 9-12) biology instruction based on a large national survey that samples science instruction at all levels. Some relevant findings from this survey:

- **Exams are commonplace**. 92% of biology teachers reported administering a quiz or test in the current unit they are teaching.
- **Biology teachers depend on instructional materials**. 84% of biology teachers—more so than that of other science disciplines—reported using commercially developed instructional materials. These tended to be (58%) single commercially-available text-

books from Pearson, Houghton Mifflin Harcourt, and McGraw-Hill. 76% of biology teachers thought their textbooks were good, very good or excellent.

- **Classroom activities skew towards lower level cognitive demand**. The most commonly reported biology classroom activities are, in order, (1) teacher explaining a science idea to the whole class, (2) whole class discussion, (3) students completing textbook/worksheet problems, and (4) students reading about science. While the Horizon study does not make any claims about the relative intellectual demand for these sorts of activities, they as a whole fall into a model of "teacher centered instruction" that dominates in high schools (Cuban, 1993).
- **There is not much support**. 40% of biology teachers indicated they had received less than 15 hours of science-focused professional development over the past three years. Given that some studies expect as much as 100 hours of professional development to see a change in student learning (S. Wilson, 2013) this is a remarkably low number.

Based on this data, the situation in Chicago Public School high school science classrooms likely is not particularly different from those around the country. What this study adds to this picture, however, is the sense of scale: For the city of Chicago, we know now the decisions teachers are making. This study provides much more local specificity to the above, and as described above, shows how homogenous the decisions about content coverage and intellectual demand are in Chicago science classrooms as well as highlights some dysfunctions in policy.

The Thought Processes Of Science Teachers

This leaves us with a somewhat frustrating picture of what teaching biology likely feels like to most teachers. Based on the evidence of the exams themselves and on the responses provided by focus group participants, teacher's mental monologue might be portrayed as follows:

"I'm teaching high school biology again this year. The district has given us a set of frameworks that describe what I should cover in this course, but they're pretty similar to what's in the textbooks at our school—and those textbooks explain what to do day-by-day much more clearly, so I'll just use them. Plus, everyone in the building has been using these same tools for years. There's not a lot of laboratory equipment or resources at our school, and the kids are pretty far behind and pretty tough, so I do as many laboratories as I can but they're hard to do. Once in a while I go downtown or to a university and they talk to us about "inquiry" and the like and I mostly understand why that would be beneficial for my students, but it's really hard to do at our school. Our principal wants us to have an end-of-semester exam, but she wants it to be "ACT style" multiple choice—thankfully they gave us a Scantron machine last year, which makes grading easy. I think the principal spends most of the time with the English and mathematics teachers—she doesn't come to my classroom much, and I'm not sure how much biology she really is comfortable with. Sometimes I try to talk to other teachers about what they're doing, but we all find it pretty hard work and don't have that much time to really collaborate on anything—it's just easier to keep my head down and move forward on my own."

<u>Why Do We Find These Things?</u>

I will now consider some possible reasons for these implications. Here, I will provide a more in depth analysis of the several gaps between policy and practice at national, state, and local levels, and analyze of the causes of the gap.

The System

Berwick is attributed with the famous line that "every system is perfectly designed to achieve the results it achieves" (Berwick, 1996). What about the design and functioning of the Chicago Public Schools is contributing to these results, and how might they be changed for the better?

Goals Vs. Outcomes: The Gap Is Huge

Does the Chicago Public Schools expect the outcomes that this research shows? Clearly, the answer is no. Over the past several decades, CPS has released numerous documents describing goals, visions, and strategies for high school teaching that certainly do not support.

The best indicator of district goals and visions at the point in time from which these exams were collected was the *Every Child, Every School* education plan. This document was released in September 2002 to "set out a out a clear vision for instruction and school development so that Chicago Public Schools is the premier urban school district in the United States" (Chicago Public Schools, 2002). It calls for quality instruction that "requires that teachers have a research-based framework for how instruction should be organized in ways that promote student learning of the content areas" and "challenging classroom assignments [that] require students to construct knowledge through interpretation and/or analysis instead of simply recalling pieces of information. Assignments ask students to draw conclusions, explain, and support their answers, and relate the assignment to their daily lives.," though this report is silent about final exams. In the district's latest strategic plan, called *The Next Generation: Chicago's Children,* the district calls for "a 21st century public education [that] must develop students who are innovative thinkers, civic-minded collaborators and effective communicators" and describes a set of supports—including common curriculum and professional development as if there has not been much change in the past decade (Chicago Public Schools, 2013).

System Breakdown

So if the intentions and goals of the district are not aligned to these outcomes, some aspects of the system must be contributing to this condition. In this section, I will consider some the likely culprits and discuss options for improvement.

Teacher Preparation

One cause for this culture of science teaching might be poor preparation of science teachers. The research base on effective STEM teacher preparation remains thin: "Despite 20 years of guidance from professionals about what teachers need to know and be able to do, there is little empirical research to support what makes particular teacher preparation, professional development, or school leadership strategies effective in improving teacher quality or student outcomes." (Community for Advancing Discovery Research in Education (CADRE) at the Education Development Center, 2012). Certainly content knowledge matters (Monk, 1994), though recent studies about the effectiveness of undergraduate instruction show that deep, conceptual understanding is frequently lacking in those classes (Committee on the Status Contributions and Future Directions of Discipline-Based Education Research, 2012), which would likely mean more challenges for teachers in producing classroom content with appropriate amounts of rigor. In Chicago, new teachers tend not to come from selective undergraduate schools (McKinsey & Company, 2001), likely indicating even less preparation. We also know that teachers frequently are not prepared for the classroom management realities of urban classrooms (Brown, 2004).

Given these factors, it is not too surprising that many teachers are not prepared to deliver the sorts of rigor and content in their classrooms that we might hope for. We do not know much about how to best prepare students to be STEM teachers in urban classrooms; in Chicago, teachers are recruited from middle- and lower- tier institutions, and even at the best undergraduate institutions STEM instruction leaves much to be desired. Efforts to improve teacher preparation are clearly needed, including extending this preparation and support to the first few years of a teacher's career in the Chicago Public Schools.

Professional Development

Another systemic factor that could be contributing to this situation is the professional development that is provided to in-service science teachers. There is an emerging literature base that describes what effective professional development for science teachers should look like (S. Wilson, 2013). Chicago Public Schools professional developments are quite scattered (Miles, Hornbeck, & Fermanich, 2002), despite efforts to reform them (Wenzel, 2010). Even well funded, comprehensive, high-profile efforts to improve professional development along principles that Wilson discusses such as the Gates-funded Instructional Development System effort met with mixed results (Sporte et al., 2009).

Absent any support for alternative methods and mechanisms for deciding upon content and rigor—which could come from a robust professional development strategy—it is likely that the traditions and cultures that define science teaching will not change. And given that most support efforts in mathematics originate external to the schoolhouse (Spillane, 2005), and science efforts are likely to follow suit, efforts to enhance the interfaces between schools and external partners are likely needed.

Curriculum and Instructional Materials

As shown in chapter 4, Chicago science teachers depend on their instructional materials to make decisions about lesson content, structure, and sequence. This matches the literature (Committee on Developing The Capacity To Select Effective Instructional Materials, 1999) and my own experience. Unfortunately, the instructional materials that are popular in Chicago are not particularly "research based" or supportive of student learning (AAAS Project 2061, 2005). Efforts to replace existing CPS instructional materials with those that are more research based led to positive feedback from teachers about these new tools, especially when delivered in a coherent set of supports (Sporte et al., 2009). Content and rigor are not likely to change in CPS science classrooms unless the instructional material situation is attended to.

School Culture and Leadership

It is clear that school culture and leadership also contribute to this problem. There is emerging evidence in the literature that "the subject matters" when issues of school and district leadership are concerned (Spillane & Hopkins, 2013). From this data, there is evidence that principals and other administrators are not particularly engaged in the instructional issues around high school science. This is not news; Richard Elmore described the "loose coupling" of schools as organizations in 1999, focusing specifically on the disconnect between school leaders and subject matter teaching and learning (Elmore, 1999). One piece of supporting evidence in this study is that many schools have multiple biology exams, indicating a lack of coherence within the school. Another is the content and intellectual demand gaps. Schools are quite rationally pushed towards the demands of the accountability system, but lack the capacity to understand or operationalize the particular demands of the assessment tool. And third, teachers feel pressure to defend their scores and grades so gravitate to the easiest measures of identifying correct and incorrect questions on their exams—and the source of much of that pressure is school leadership. Better strategies to develop and support school leaders and school cultures in the areas of science would likely help (R. Halverson, Feinstein, & Neshoulam, 2011).

The Routine Of Final Exams

One of the big takeaways here is that the routine of the end-of-semester final exam is deeply ingrained into the culture and tradition of the Chicago Public Schools—even nationally, as the aforementioned references indicate. But it is not accomplishing what it intends to. There has been relatively little attention paid to this routine—there are no policies about final exams, little guidance to teachers or schools about what to do about them, and little direction given to teachers. While the Chicago Public Schools have made some attempts at centralized end-of-course exams, namely the Chicago Academic Standards Exams (CASE) (G. N. Schmidt, 2007) and the IDS exams developed by American Institutes for Research (Sporte et al., 2009) currently neither of these are used and there is nothing taking their place on the district's assessment calendar (see http://www.cps.edu/Performance/Documents/CPSSY14AssessmentCalendar.pdf). Given the attention placed on this routine by teachers and schools, it is a space that is ripe for some definition or attention from the central office or other organizations.

Implications For Practice

So what is to be done? There are several implications for practice that emerge from this.

New Standards

It is clear that advocates of the *Next Generation Science Standards* have their work cut out for them. Getting states to adopt the new standards is a big challenge. But since standards interact deeply with all other aspects of the educational system—including curricula, assessments, human capital, school organization—isolating collateral impacts is difficult. Standards alone accomplish very little (C-STEMEC, 2013). The introduction of the NGSS to a school, district, or state should initiate changes in many other connected aspects of policy and practice—though we know from prior experience that the changes are often not what were intended (Spillane, 2004). This data shows both how much work needs to be done before the vision of the standards is manifest in science classrooms around the country, but also the incredible interconnectedness of leadership, accountability, resources, and culture that standards implementation plans will need to tackle. Repeating this study in a few years, after NGSS have been adopted and "rolled out," would be a powerful way to measure their impact.

Tending To The System

The findings here suggest a culture of science teaching that pervades the system—it is widespread and homogenous, and not particularly different in Chicago or the nation. In *The Teaching Gap,* Stigler and Hiebert argue that many educational reforms fail because they do not attend to the particular cultures of teaching (Stigler & Hiebert, 1999). If we want systemic efforts to succeed, they will need to address these cultural factors of teaching. Clearly, the current implementation of the final exam routine contributes to the culture of teaching for science teachers.

Given the homogeneity observed in the data here, it appears unlikely that recent efforts to promote autonomy (Hill, 2013) among school systems will reap great benefits for science learning. The data presented here is both homogeneous and quite broad, indicating that variance in content coverage and intellectual demand is small despite a large sample. Without greater differences, it will be difficult to engender a community or system to drive improvements—if everyone has essentially the same practice, more freedoms will not create change. (My experience tells me that teachers in the Chicago Public Schools already have a large degree of autonomy over classroom decisions.) The culture of content coverage and low intellectual demand will suffocate well meaning and innovative efforts that at best, will not get beyond the hothouse (Committee on High School Science Laboratories: Role and Vision, 2005, page 87). Instead, efforts should focus on coherently attending to the myriad of systemic issues, and injecting new capacity into the system from external sources—such as museums, curriculum developers, scientists, and researchers (Tobias & Baffert, 2009). Clarity about the goals and purposes of final exams would certainly help. As evidence suggests that the decisions of school district leaders are "are shaped in crucial ways by preexisting working knowledge and practices that guide how people come to understand the nature of problems and possible avenues for solutions" (Coburn, Touré, & Yamashita, 2009) support from outside entities is essential, particularly since the subject seems to matter for district as well as school level leadership (Burch & Spillane, 2005). Efforts to tackle this culture head-on—by bringing transparency to end-of-course exams that are teacher generated, for instance—are worth considering.

New Lessons

Several new lessons emerge from these implications that can help with interim steps towards reform.

Exams could be a useful proxy for instruction, particularly if the final exam routine is also improved. Examining teacher generated end-of-semester exams as a proxy for instruction within classrooms is a useful exercise that should continue to yield important information. Given that "task predicts performance," an analysis of the content and intellectual demand that teachers offer to students as their summative measure can shed light on many aspects of practice. If established as a regular routine in a school or district, it has the potential to be a cost-effective and illuminating evaluation tool or a beneficial professional development endeavor. Even assuming higher costs than this study took to complete, such as $100 per exam-review, 2 reviewers per exam, 1 common biology exam per school, this is less expensive than video analysis or classroom observations to do. (Particularly if the laborious collecting, organizing, and tracking processes can be attended to by the system.). This is a cost effective measure to analyze practice at scale—in this case, the whole district—particularly if the purposes of final exams are clearly defined. When considered broadly, the analysis of these exams presents a picture that is remarkably similar to that of the nation—but from a very different data source, and with considerably more nuance than grades or standardized tests currently provide. It enables school comparisons at scale, which is potentially powerful for reform efforts. Moving forward, systems initiated by schools and districts to highlight assessment practices—particularly when connected to issues of content and intellectual demand—have the potential to focus attention on key instructional issues (Kanold, 2003). Moreover, schools and districts could use the regular sharing and reporting of end-of-semester exams as a tool to gauge the effectiveness of instructional interventions. It enables us to tend to the instructional core at scale in a new, potentially powerful, fashion.

Tools matter. The tools that science teachers have—textbooks, curriculum frameworks, pacing guides—largely define what is valued by and large by teachers. As we saw here, the content coverage of Chicago Public School end-of-semester exams largely mapped to the vestigial pacing guides the district shared, which appear to be built off the most popular (traditional) textbooks that schools have purchased. Given that significant differences emerge when considering leadership (Spillane, 2005) and instruction (Graeber, Newton, & Chambliss, 2012) in different disciplines, and the usage of these tools appears different in science than in other disciplines (Wenzel, 2010), this speaks to the need for schools and districts to invest in content-focused science tools and leadership to help teachers and schools make better decisions in these areas.

Leaders need to pay attention to science. The data presented here should not be particularly encouraging to school and district leaders. A common theme, however, is not only that school and district leadership matters, but that the content matters as well—if we want teachers to produce exams that better address the content and intellectual demand we expect for students, leaders need to be comfortable enough with the content and context to lead in these areas. The much-ballyhooed role of "principals as instructional leaders" means little if principals do not understand how to lead instruction in specific subject matter domains. This does not mean that principals need to be experts in every subject matter; it does mean that their leadership must attend to the distinctive character of teaching and learning in each subject area. Elmore discussed this at length in *School Reform from the Inside Out,* which focused on the "instructional core" of the relationships among teacher, student, and content (Elmore, 2004).

Change is slow. There was not much to say the content of biology changed significantly from many years ago—though we know as a field, the content changes dramatically. District pacing guides and basal-focused textbooks still get tremendous attention from teachers and classrooms. The more school and district leaders can give attention and cover to teachers and schools who are actively working on these changes, the better.

The Chicago Public Schools context is challenging. The overall sentiments of the teacher focus group were not particularly encouraging. Teachers commented and complained about the low level of their students, the incredible demands on their time, and the lack of resources that were available to them. We should not forget just how challenging the work of teaching, learning, and leading science education is, particularly in districts as historically under resourced and systemically challenged as Chicago.

Lastly, and most fundamentally, in reviewing the data and thinking back on the teacher focus group conversations, I am struck by the similarities between these findings and the sobering picture of the American high school painted many years ago by Ted Sizer in the classic *Horace's Compromise* (Sizer, 1992). In that book, Sizer describes the compromise that teachers make everyday with their students, their schools, and themselves: the work is too hard, and the resources are spread so thin, so teachers end up not pushing anyone too hard for what they know is right. Instead, they go through the motions, pretending to teach their students as they in turn pretend to learn. It is a vicious cycle, but one that Chicago Public School science teachers are clearly in the throes of. They can not shake it—or at least in the several-year period in which the documentary and focus group data were obtained for this study, there were no signs of significant change.

Implications For Research

This study opens the door to plenty of additional research that could be done.

New tools for new standards. With the arrival of the *Next Generation Science Standards,* the potential for states to work together on issues of curriculum is greater than ever before. This means more collective and coordinated resources and more attention than the science education community has seen in a long time. These new standards will require new tools for teachers—new assessments, new curriculum, and new ways to organizing and thinking about the inherent content and intellectual demand that is part of it.

End-of-semester exams as proxies for other measures. This research demonstrates the potential of using teacher generated assessments as proxies for instructional and achievement progress at schools. Because these exams are a common routine in classrooms, relatively easy and inexpensive to analyze at scale, and filled with lots of information that describe what teachers believe is important for students to know and be able to do, they are a strong proxy to understand "how teachers are teaching". Larry Cuban comments about the challenges of determining what sorts of instructional practices were used in classrooms:

> *"…Records of classroom lessons or observations are rarely available to researchers. Moreover, interviewing or surveying teachers [about] how they taught… often yields unreliable results. For example, surveys of teachers, the most common and least expensive way of ascertaining classroom practices, remain imprecise and tend to reflect what teachers believe they did, not what occurred when independent observers sat in their rooms." (Cuban, 2007)*

Efforts that explore the nature of the relationship between teacher-generated assessment tasks and other factors are worth exploring.

Assessments To Grades. This study ends at teacher generated assessments, but as we all know, assessments are in turn connected to student grades. More data about those connections would be welcome. For instance, how did the mix of items on end-of-semester exams get translated into an exam grade? How did those exam grades impact the final semester grade? We know that there exists considerable variation in high school grading scales across the district (The Chicago Sun Times, 2009), and that course grades matter a great deal to student overall success (E. M. Allensworth & Easton, 2007), and seem to be based on student achievement (Randall & George Engelhard, 2009), so exploring more how teachers use the exams they generate to create student end-of-course grades would be very illuminating.

More nuanced content. As mentioned in chapter 5, this analysis scratches the surface of the data that has been collected. In the *Survey of Enacted Curriculum* coding specifications, a much finer grain of content alignment is possible—and has been generated for these exams via this coding process. With this data, for instance, it should be possible to delve into more detail about content particulars—what aspects of cell biology are most commonly assessed, or how to teachers construct ideas about the kinetic molecular model of matter, for instance. We might find that the observed homogeneity in content is less obvious when more detailed content criteria are used. That analysis remains to be done.

Benchmarking reforms. Lastly, a study like this could provide a relatively cost effective way to benchmark district progress in science if conducted every few years. Particularly if the district assumed the burden of collecting and end-of-semester exams, a third party research group could analyze them in a few months for relatively little cost—an easy way to tell if curricular, professional development, or other support efforts were taking root.

Conclusions

Tools that enable us as both researchers and practitioners to explore this technical core at scale have the potential to be quite powerful. In this study, I used teacher-generated end-of-semester final exams to investigate the depth and breadth of content that students learn in science classrooms. I found that teachers focus on a specific set of scientific content that is driven by district guidelines and popular textbooks but not particularly aligned to standards. Rigor means coverage instead of intellectual press. The assessments, while unsophisticated, seem to be delivering what is expected of them—a way to mimic the format of the ACT exam quickly. And there was little variation among high poverty and low poverty schools, matching national data and indicating issues that are more due to a particular culture of science teaching and learning than driven by particular contexts.

Appendix 1: *Survey Of Enacted Curriculum* Content Codes

K-12 Science Content Areas

100 **Nature of Science**	1500 **Motion & Forces**
200 **Science & Technology**	1600 **Electricity**
300 **Science, Health & Environment**	1700 **Waves**
400 **Measurement & Calculation in Science**	1800 **Kinetics and Equilibrium**
500 **Components of Living Systems**	1900 **Properties of Matter**
600 **Biochemistry**	2000 **Earth Systems**
700 **Botany**	2100 **Astronomy**
800 **Animal Biology**	2200 **Meteorology**
900 **Human Biology**	2300 **Elements & The Periodic System**
1000 **Genetics**	2400 **Chemical Formulas & Reactions**
1100 **Evolution**	2500 **Acids, Bases, & Salts**
1200 **Reproduction & Development**	2600 **Organic Chemistry**
1300 **Ecology**	2700 **Nuclear Chemistry**
1400 **Energy**	

K-12 SCIENCE TAXONOMY August 2004

100	**Nature of Science**
101	Nature and Structure of Science
102	Nature of Scientific Inquiry/Method
103	Scientific habits of mind, logic and reasoning
104	Issues of diversity, culture, gender in science
105	History of scientific innovations
106	Ethical Issues/Critiques of Science
190	Other
200	**Science & Technology**
201	Tech. benefits, trade-offs and consequences
202	Relationship btwn. sci. inquiry & tech. design
203	Science tools, lab safety
204	Design or implement a solution or product
290	Other
300	**Science, Health & Environment**
301	Personal health, behavior, disease, nutrition
302	Envrn. health, pollution, waste disposal
303	Acid rain
304	Ozone depletion
305	Resources, conservation
306	Toxic & nuclear waste
307	Greenhouse effect
308	Natural and Human-caused hazards
390	Other
400	**Measurement & Calculation in Science**
401	The International System
402	Mass & Weight
403	Length
404	Volume
405	Time
406	Temperature
407	Accuracy & Precision/Estimation
408	Significant Digits
409	Derived Units
410	Conversion Factors
411	Density
490	Other
500	**Components of Living Systems**
501	Cell structure/function
502	Cell Theory
503	Transport of cellular material
504	Cell metabolism
505	Cell response
506	Cellular respiration
507	Cell Specialization
508	Organs
509	Organ Systems
510	Microbiology
590	Other
600	**Biochemistry**
601	Living Elements (C, H, O, N, P)
602	Atomic Structure & Bonding
603	Synthesis Reactions (Proteins)
604	Hydrolysis
605	Organic Compounds: Carbon, Proteins, Nucleic/Amino Acids, Enzymes
690	Other
700	**Botany**
701	Nutrition/Photosynthesis
702	Circulation
703	Respiration
704	Growth/development/behavior
705	Health & disease
706	Structure & Function
790	Other
800	**Animal Biology**
801	Nutrition
802	Circulation
803	Excretion
804	Respiration
805	Growth/development/behavior
806	Health & disease
807	Structure & Function
808	Skeletal & muscular system
809	Nervous & endocrine system
810	Habitat
890	Other
900	**Human Biology**
901	Nutrition/Digestive System
902	Circulatory System (Blood)
903	Excretory System
904	Respiration & Respiratory System
905	Growth/development/behavior
906	Health & disease/immune system
907	Skeletal & muscular system
908	Nervous & endocrine system
990	Other
1000	**Genetics**
1001	Mendelian Genetics
1002	Modern Genetics
1003	Inherited diseases
1004	Biotechnology
1005	Human Genetics
1006	Transcription/translation
1007	Mutation
1090	Other
1100	**Evolution**
1101	Evidence for Evolution
1102	Lamarckian Theories
1103	Modern Evolutionary Theory
1104	Life Origin Theories
1105	Human Evolution
1106	Classification
1107	Causes
1108	Natural Selection
1190	Other
1200	**Reproduction & Development**
1201	Mitotic/Meiotic Cell Division
1202	Asexual Reproduction
1203	Inherited Traits
1204	Reproduction & Development in Plants
1205	Reproduction & Development in Animals
1206	Reproduction & Development in Humans
1290	Other

K-12 SCIENCE TAXONOMY August 2004

1300	**Ecology**
1301	Food Webs / Chains
1302	Competition & Cooperation
1303	Energy Flow Relationships
1304	Biotic & Abiotic Factors
1305	Ecological Succession
1306	Ecosystems
1307	Population Dynamics
1308	Environmental Chemistry
1309	Adaptation & Variation / Niche
1310	Populations
1390	Other
1400	**Energy**
1401	Potential Energy
1402	Kinetic Energy
1403	Conservation of Mass/Energy
1404	Heat Energy & Transfer
1405	Light Energy
1406	Sound Energy
1407	Laws of thermodynamics & entropy
1408	Work & Energy
1409	Mechanical Energy & Machines
1410	Nuclear Energy
1490	Other
1500	**Motion & Forces**
1501	Vector & Scalar Quantities
1502	Displacement as a vector quantity
1503	Velocity as a vector quantity
1504	Relative position & velocity
1505	Acceleration
1506	Newton's First Law
1507	Newton's Second Law
1508	Newton's Third Law
1509	Momentum, Impulse and Conservation
1510	Equilibrium
1511	Friction
1512	Universal Gravitation
1590	Other
1600	**Electricity**
1601	Static Electr.(production/transfer/distribution)
1602	Coulomb's law
1603	Electric fields
1604	Current electricity
1605	Current, Voltage, & Resistance
1606	Series & Parallel Circuits
1607	Magnetism
1608	Effects of interacting fields
1609	Conductors, insulators
1690	Other
1700	**Waves**
1701	Characteristics and behavior
1702	Visible Light (direction/speed/transformation)
1703	Non-visible Light/Electromagnetic Spectrum (e.g. ultraviolet, infrared)
1704	Sound (e.g. direction, speed, transformation)
1705	Earthquakes, Tsunamis, Ocean Waves
1790	Other

1800	**Kinetics and Equilibrium**
1801	Molecular motion
1802	Pressure
1803	Kinetics and temperature
1804	Equilibrium
1805	Reaction Rates
1890	Other
1900	**Properties of Matter**
1901	Characteristics & composition
1902	Elements, molecules & compounds
1903	States of matter (S-L-G-P)
1904	Solutions & Mixtures
1905	Physical & Chemical Changes
1906	Physical & Chemical Properties
1907	Isotopes/Atomic Nbr./Atomic Mass
1908	Photons & Spectra
1909	Atomic Theory
1910	Quantum Theory & Electron Clouds
1990	Other
2000	**Earth Systems**
2001	Earth's shape, dimension & composition
2002	Earth's origins and history
2003	Maps, locations and scales
2004	Measuring using relative and absolute time
2005	Mineral & Rock Formations & Types
2006	Erosion & Weathering
2007	Plate Tectonics
2008	Formation of volcanoes, earthquakes, mtns.
2009	Topography
2010	Dynamics & Energy Transfer
2011	Oceanography
2090	Other
2100	**Astronomy**
2101	Stars, Sun
2102	Galaxies
2103	Origins of the universe
2104	Asteroids and comets
2105	The Solar System
2106	The Moon
2107	The Earth's motion: rotation & revolution
2108	Earth, moon, sun relationship
2109	Location, Navigation, & Time
2190	Other
2200	**Meteorology**
2201	The Earth's Atmosphere
2202	Air Pressure & Winds
2203	Evaporation / Condensation / Precipitation
2204	Weather
2205	Climate
2290	Other
2300	**Elements & The Periodic System**
2301	Early Classification System(s)
2302	Modern Periodic Table
2303	Interaction of elements
2304	Element char. (families & periods)
2390	Other

2400	**Chemical Formulas & Reactions**
2401	Names, Symbols, & Formulas
2402	Molecular & Empirical formulas
2403	Representing chemical change
2404	Balancing chemical equations
2405	Stoichiometric Relationships
2406	Oxidation/Reduction Reactions
2407	Chemical Bonds
2408	Electrochemistry
2409	The Mole
2410	Types of reactions
2490	Other
2500	**Acids, Bases, & Salts**
2501	Arrhenius/Bronsted-Lowry/Lewis Theories
2502	Naming Acids
2503	Acid-Base behavior/strengths
2504	Salts
2505	pH
2506	Hydrolysis
2507	Buffers
2508	Indicators
2509	Titration
2590	Other
2600	**Organic Chemistry**
2601	Hydrocarbons, Alkenes, Alkanes, & Alkynes
2602	Aromatic Hydrocarbons
2603	Isomers & Polymers
2604	Aldehydes, Ether, Ketones, Esters, Alcohols, & Organic Acids
2605	Organic Reactions
2606	Carbohydrates, Proteins, Lipids
2690	Other
2700	**Nuclear Chemistry**
2701	Nuclear Structure
2702	Nuclear Equations
2703	Fission
2704	Radioactivity
2705	Half-life
2706	Fusion
2790	Other

Appendix 2: Focus Group Informed Consent Form

University of Illinois at Chicago
Consent for Participation in Research
"High School Science Semester Tests In The Chicago Public Schools"

Why am I being asked?

You are being asked to be a subject in a research study about high school science exams in the Chicago Public Schools conducted by Michael Lach of the College of Education at the University of Illinois at Chicago. You have been asked to participate in the research because you are a high school science teacher and may be eligible to participate. We ask that you read this form and ask any questions you may have before agreeing to be in the research.

Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University or the Chicago Public Schools. If you decide to participate, you are free to withdraw at any time without affecting that relationship.

Why is this research being done?

In science instruction, breadth and depth of learning both matter. As new scientific discoveries are made on a daily basis, the overall breadth of scientific knowledge is increasing continually. Teachers and school leaders know this must be moderated in the classroom—an introductory biology course can't focus solely on genetics, for instance. At the same time, students need to learn both the unique ways of knowing associated with the scientific endeavor as well as sufficient content to transcend the often-naïve misconceptions that are developed in their everyday interactions with the natural world. This means studying particular topics with depth.

The rise of standards-based reform over the last few decades in American education created a wave of cascading policies and changes. Yet the heart of this work—the teaching and learning at the technical core—remains "loosely coupled" from classrooms, despite all the attention and energy that have focused upon it.

Tools that enable us as both researchers and practitioners to explore this technical core at scale thus are potentially powerful. To investigate the depth and breadth of con-tent that students learn in science classrooms, one tool is the summative exams that teachers write and students complete. End-of-semester and end-of-course exams are hallmarks of the American high school, and ostensibly address both the depth and breadth of content contained within a course. And while individual teacher-written exams undoubtedly vary greatly, analysis of large collections of such exams can reveal patters than describe the depth and breadth of science learning expected of Chicago high school students.

In the 2003-04 school year, several hundred final exams from high school science classrooms were collected as part of a school improvement strategy initiated by the Chicago Math & Science Initiative. By studying these actual artifacts of individual classroom practice, this research will explore an important part of the technical core of high school science teaching and learning. The final product will be a picture of classroom practice that explains what high school science teachers in Chicago value about the content they teach, how they organize that content, and the sort of demands they expect of their students.

What is the purpose of this research?

This research attempts to answer three questions.

1. **Content**: What content is valued by Chicago high school science teachers in the courses they teach, as reflected in the assessments they use? How does this content relate to state and national standards? How are topics with important reform or societal connotations (e.g. evolution, the environment, inquiry, nature of science) treated?

2. **Rigor**: What is the level of "academic press" associated with Chicago high school science courses, as reflected in the assessments used in the courses they teach? How does this level vary within and between schools?

3. **Assessment Literacy**: What does local assessment design, particularly on final exams, say about the level of assessment literacy of CPS high school science teachers?

What procedures are involved?

If you agree to be in this research, we would ask you to participate in a focus group with fellow high school science teachers. The purpose of this focus group is to probe the decision making processes of teachers around the topic of semester exam construction. A list of focus group questions is attached to this document. The focus group is scheduled to take about 2 hours, and will be facilitated by Michael Lach, the investigator of this study. The focus group will be tape recorded, and the discussion will be transcribed. The transcript of the focus group, with all unique identifying information removed, will then be circulated to the focus group participants for review.

Approximately 20 current and former CPS high school science teachers may be involved in this research at the University of Illinois at Chicago.

What are the potential risks and discomforts?

There are several potential risks associated with participating in this research.

First, as part of the focus group, you will be asked to complete and sign a document indicating you are providing informed consent. These documents will be stored securely at the home of the researcher.

Second, as part of the focus group, you will be asked to provide insight and information into issues associated with teaching and learning science in the Chicago Public Schools. The focus group will be recorded and notes will be transcribed. These documents will be stored securely at the home of the researcher.

Third, the focus group will be conducted by the investigator, who is a member of the senior staff of the Chicago Public Schools. The focus group work will not be substantially differrent from conversations the investigator has with teachers on a daily basis. The investigator has no authority over any in-school teaching positions within the Chicago PUblic Schools. Furthermore, the role of the focus group is to confirm, disconfirm, and augment my interpretations of findings that come from the analysis of the course exams, not to generate new avenues for study or inquiry.

Are there benefits to taking part in the research?

By participating in this research, you'll help advance the understanding of science education in practice with the Chicago Public Schools. This will help other teachers, university faculty, school leaders, and policy makers enhance the tools and supports for teachers and students in the future. Your participation will also connect you to other high school science teachers within CPS who may have insights that will help you in your practice.

What about privacy and confidentiality?

The only people who will know that you are a research subject are members of the research team. No information about you, or provided by you during the research, will be disclosed to others without your written permission, except:

- if necessary to protect your rights or welfare (for example, when the UIC Institutional Review Board monitors the research or consent process); or
- if required by law.

When the results of the research are published or discussed in conferences, no information will be included that would reveal your identity.

Any information that is obtained in connection with this study and that can be identified with you will remain confidential and will be disclosed only with your permission or as required by law.

A tape recording of the focus group you are being asked to participate in will be made. This tape recording will be sent to an external party to be transcribed. The transcript will be reviewed by the focus group members for accuracy and completeness..Once the transcript has been provided to the investigator, all notes, tapes, and recordings of the focus group session will be destroyed.

It is important to note that within a focus group setting, it is important to respect the confidentiality and privacy of other focus group participants. Please do not discuss or share questions or responses or insights from the focus group with anyone. Because of the nature of focus group conversations, the investigator cannot guarantee privacy or confidentiality in this setting.

What are the costs for participating in this research?

There are no costs for participating in this research.

Will I be reimbursed for any of my expenses or paid for my participation in this research?

Each focus group participant will receive a $25 gift certificate for Amazon.com for participating in the focus group.

Can I withdraw or be removed from the study?

You can choose whether to be in this study or not. If you volunteer to be in this study, you may withdraw at any time without consequences of any kind. You may also refuse to answer any questions you don't want to answer and still remain in the study. The investigator may withdraw you from this research if circumstances arise which warrant doing so.

\\

Who should I contact if I have questions?

The investigator conducting this study is Michael Lach. You may ask any questions you have now. If you have questions later, you may contact the him at 773/841-0149 or mlach@teachandlearn.org.

The investigator is a student in the UIC Urban Education Leadership program. His advisor is Steve Tozer, (312) 413-7782, stozer@uic.edu.

What are my rights as a research subject?

If you feel you have not been treated according to the descriptions in this form, or you have any questions about your rights as a research subject, you may call the Office for the Protection of Research Subjects (OPRS) at 312-996-1711 (local) or 1-866-789-6215 (toll-free) or e-mail OPRS at uicirb@uic.edu.

Remember: Your participation in this research is voluntary. Your decision whether or not to participate will not affect your current or future relations with the University. If you decide to participate, you are free to withdraw at any time without affecting that relationship. You will be given a copy of this form for your information and to keep for your records.

Signature of Subject or Legally Authorized Representative

I have read (or someone has read to me) the above information. I have been given an opportunity to ask questions and my questions have been answered to my satisfaction. I agree to participate in this research. I have been given a copy of this form.

______________________________ ____________
Signature Date

____
Printed Name

______________________________ ____________
Signature of Researcher Date (must be same as subject's)

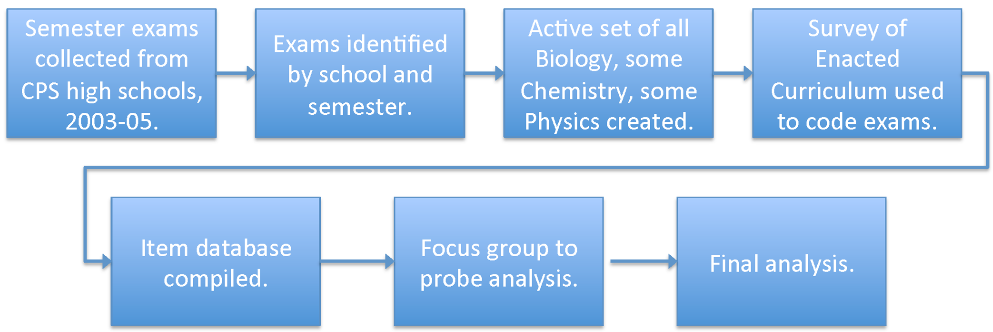
Appendix 3: Focus Group Discussion Deck

Focus Group Discussion Deck

May 19, 2011

Confidential Draft

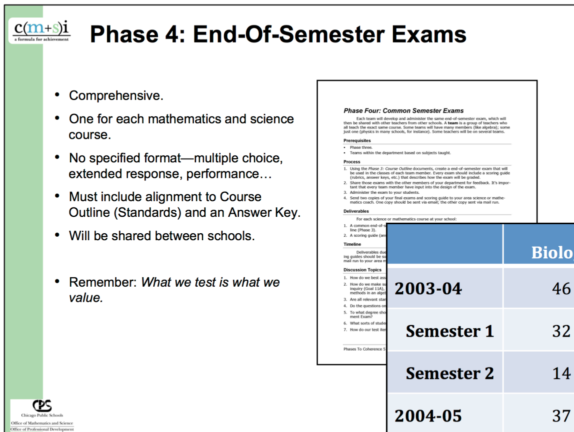
Data collection and analysis process created to answer three key questions.



1. **Content**: What content is valued by Chicago high school science teachers in the courses they teach, as reflected in the assessments they use?
2. **Rigor**: What is the level of "academic press" associated with Chicago high school science courses, as reflected in the assessments used in the courses they teach? How does this level vary within and between schools?
3. **Assessment Literacy**: What does local assessment design, particularly on final exams, say about the level of assessment literacy of CPS high school science teachers?

Confidential Draft

Data was collected over 2 years and a variety of CPS high schools.



Fall 2003 request to HS chairs

	Biology	Chemistry	Physics	Grand Total
2003-04	46	8	11	65
Semester 1	32	3	8	43
Semester 2	14	5	3	22
2004-05	37	3	4	44
Semester 1	34	3	4	41
Semester 2	3	0	0	3
Total	83	11	15	109

Six different item types were recorded.

Item Format	Notes
Constructed response	Any question demanding more writing than a "short answer" question, typically at least a full sentence answer, and sometimes, many paragraphs.
Fill in the blank	Students are asked to identify a missing word or phrase from a given sentence fragment.
Multiple choice	Students are asked to choose the correct response from a number of choices given. Questions that were matching (e.g. match the label on the diagram of a cell with the proper scientific term) were also considered multiple choice. Reviewers recorded the number of possible options for each item as well (*e.g.* "multiple choice – 4" or "multiple choice – 5").
Rewrite sentence	Students were asked to rewrite the given sentence—a slightly more complicated version of fill-in-the-blank.
Short answer	Students were asked to respond with the results of a calculation or a short, 1-3 word answer.
True or false	Students were asked if a statement was true or false.

Content was coded using the Survey Of Enacted Curriculum categories.

K-12 Science Content Areas

100 Nature of Science	1500 Motion & Forces
200 Science & Technology	1600 Electricity
300 Science, Health & Environment	1700 Waves
400 Measurement & Calculation in Science	1800 Kinetics and Equilibrium
500 Components of Living Systems	1900 Properties of Matter
600 Biochemistry	2000 Earth Systems
700 Botany	2100 Astronomy
800 Animal Biology	2200 Meteorology
900 Human Biology	2300 Elements & The Periodic System
1000 Genetics	2400 Chemical Formulas & Reactions
1100 Evolution	2500 Acids, Bases, & Salts
1200 Reproduction & Development	2600 Organic Chemistry
1300 Ecology	2700 Nuclear Chemistry
1400 Energy	

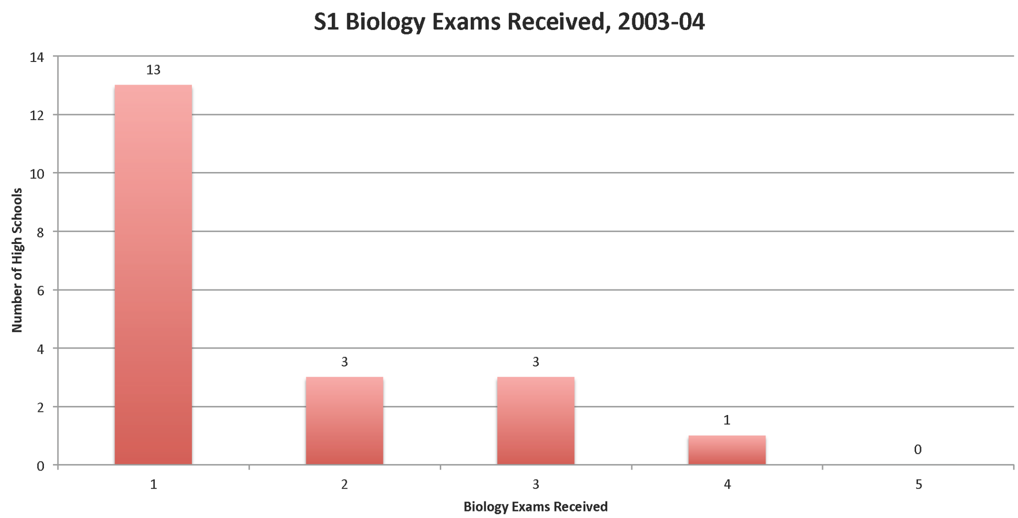
Confidential Draft

Cognitive demand was recorded using the Survey of Enacted Curriculum categories.

Cognitive Demand Categories for Science

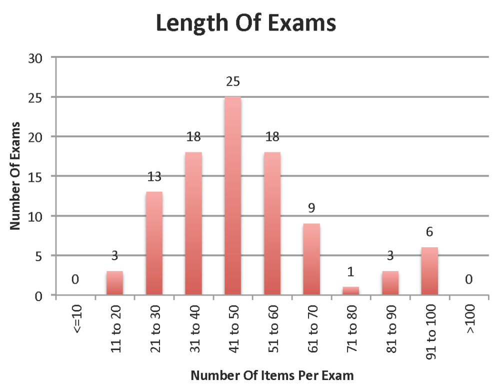
B	C	D	E	F
Memorize	Perform Procedures	Communicate Understanding	Analyze Information	Apply Concepts / Make Connections
Recite basic science facts	Make observations Collect and record data	Explain concepts	Classify and compare data	Apply and adapt science information to real-world situations
Recall science terms and definitions	Use appropriate tools Make measurements, do computations	Observe & explain teacher/student demonstrations	Analyze data, recognize patterns	Apply science ideas outside the context of science
Recall scientific formula	Organize and display data in tables or charts	Explain procedures & methods of science & inquiry	Reason inductively or deductively	Build or revise theory/plan and design experiments
	Execute procedures Conduct experiments	Organize & display data in tables or charts	Draw conclusions	Synthesize content and ideas from several sources
	Generate questions, make predictions	Student presentations of science information	Identify faulty arguments or misrepresentations of data	Use and integrate science concepts
	Test effects of different variables			

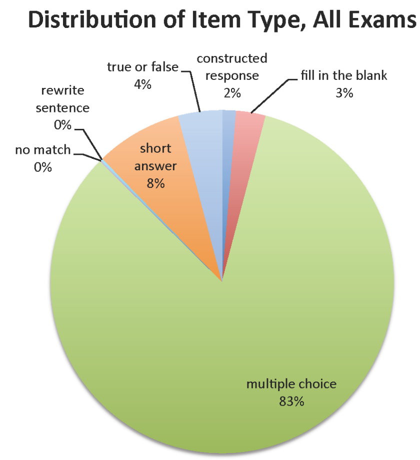
Many schools submitted several exams.



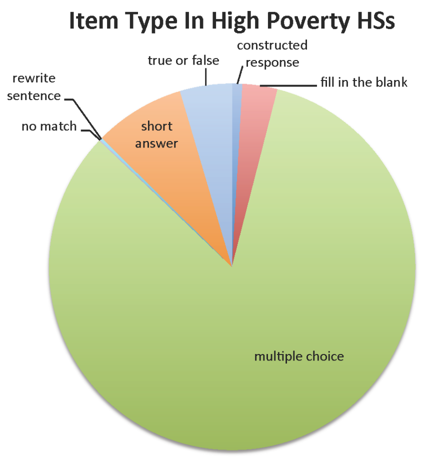
Confidential Draft

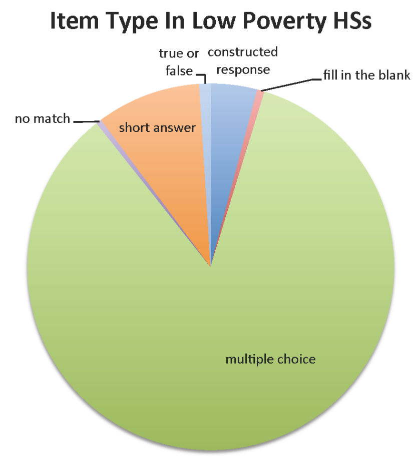
The typical exam is between 41 and 50 items, mostly multiple choice.





Item type distribution is roughly the same regardless of school poverty level.

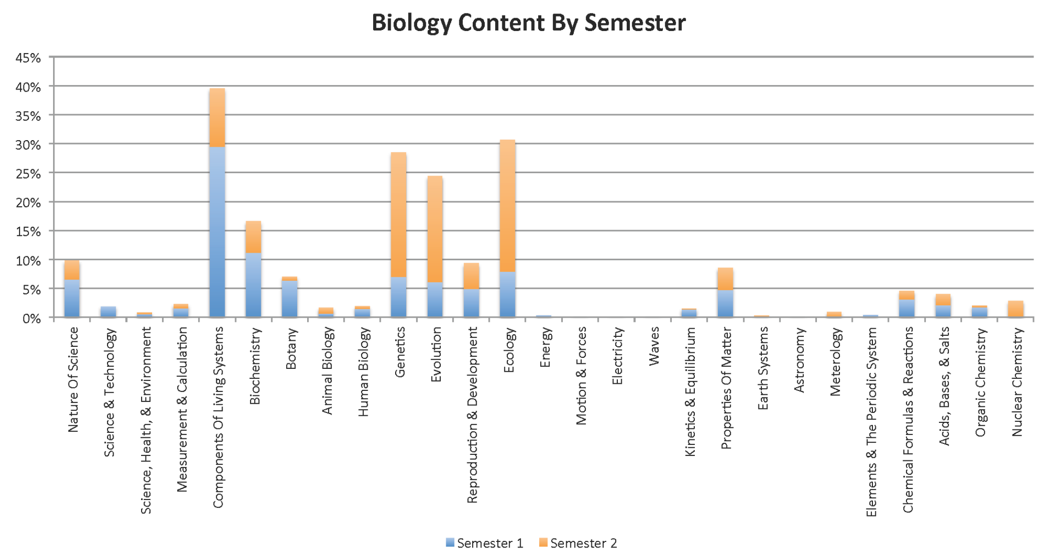




Confidential Draft

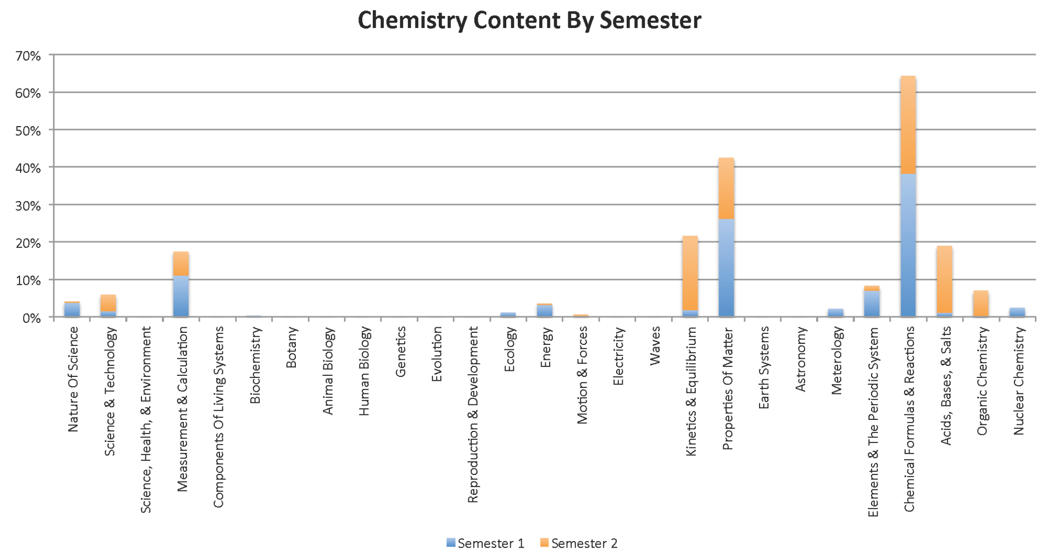
9

Biology content patterns exist at the semester level.

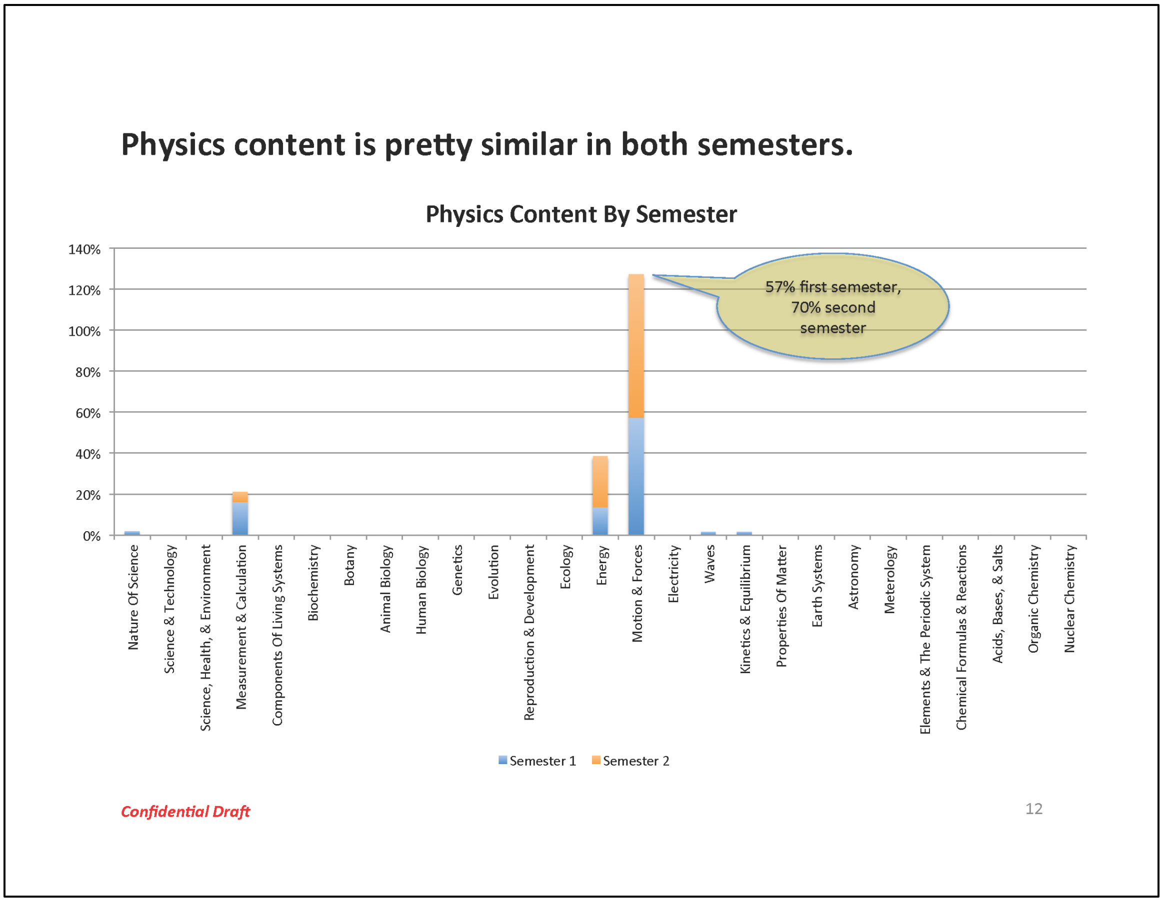


Confidential Draft

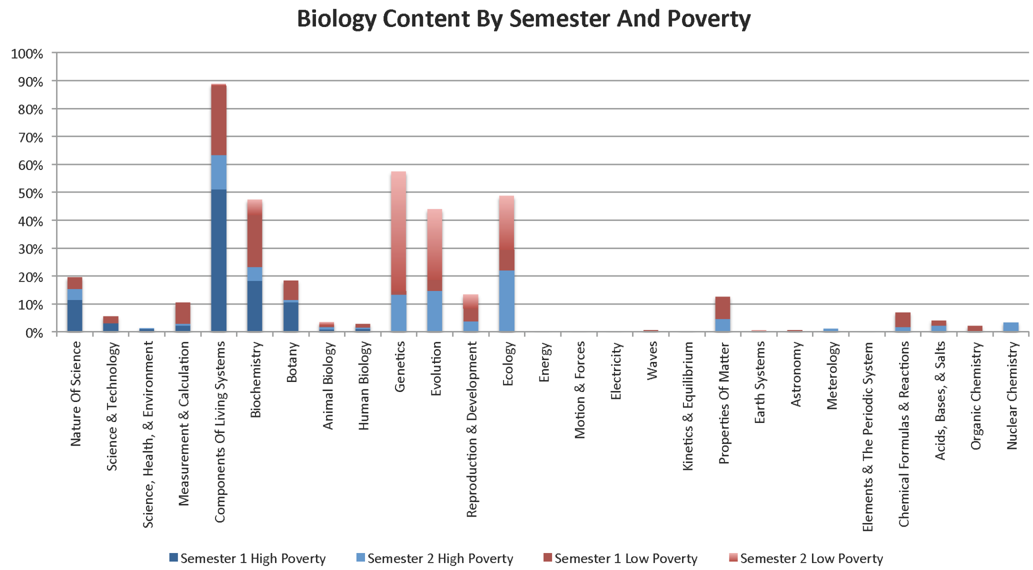
Chemistry content patterns exist at the semester level.



Confidential Draft


Physics content is pretty similar in both semesters.
Physics Content By Semester
140%
120%
100%
80%
60%
40%
20%
0%
57% first semester, 70% second semester
Nature Of Science
Science & Technology
Science, Health, & Environment
Measurement & Calculation
Components Of Living Systems
Biochemistry
Botany
Animal Biology
Human Biology
Genetics
Evolution
Reproduction & Development
Ecology
Energy
Motion & Forces
Electricity
Waves
Kinetics & Equilibrium
Properties Of Matter
Earth Systems
Astronomy
Meterology
Elements & The Periodic System
Chemical Formulas & Reactions
Acids, Bases, & Salts
Organic Chemistry
Nuclear Chemistry
Semester 1
Semester 2
Confidential Draft
12

Biology content differs slightly between high poverty and low poverty schools.



Confidential Draft

13

Biology exams have generally lower cognitive demand than other subjects.

Cognitive Demand	Biology	Chemistry	Physics
Memorize	77%	37%	44%
Perform Procedures	12%	54%	38%
Communicate Understanding	4%	1%	6%
Analyze Information	6%	8%	8%
Apply Concepts, Make Connections	1%	0%	3%
no match	1%	0%	0%
no score	0%	0%	1%

On biology exams, high poverty schools have less cognitive demand.

Cognitive Demand	low poverty	high poverty
Memorize	58%	80%
Perform Procedures	20%	10%
Communicate Understanding	5%	4%
Analyze Information	11%	5%
Apply Concepts, Make Connections	0%	1%
no match	3%	0%
no score	2%	0%

Confidential Draft

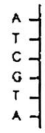
Appendix 4: Sample Exam With High Proportion Of High Cognitive Demand Items

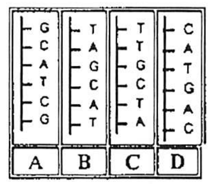
Final Exam Biology 2003-2004

[redacted] Academy

Multiple Choice: Choose the MOST correct response for the questions below. Circle the letter of the correct answer.

1. Which DNA sequence represents the complimentary base pair to the portion of DNA strand here?

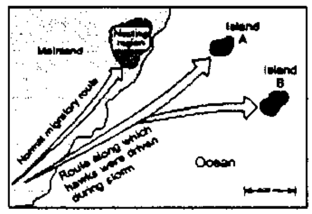




2. The DNA sequences found in two different species are 95% the same. This similarity suggests that these species
   a. Have ALL of the same proteins
   b. Have similar evolutionary histories
   c. Are the same species
   d. Are not related at all

3. According to modern evolutionary theory, genes responsible for new traits that help a species survive in a particular environment will usually
   a. not change in frequency
   b. decrease slowly in frequency
   c. increase in frequency
   d. decrease rapidly in frequency

4. Thousands of years ago, a large flock of hawks was driven from its normal migratory route by a severe storm. The birds scattered and found shelter on two distant islands, as shown on the map below. The environment of island A is very similar to the hawks' original nesting region. The environment of island B is very different from that of island A. the hawks have survived on these islands to the present day with no migration between populations.

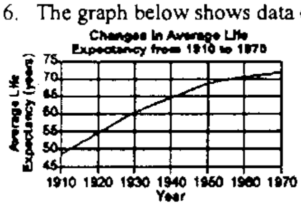


Which statement most accurately predicts the present day condition of these island hawk populations?

A. The Hawks that landed on island B have evolved more than those on island A
B. The Hawks that landed on island A have evolved more than those on island B
C. The populations on islands A and B have undergone identical mutations
D. The Hawks on island A have given rise to many new species

5. Several white potato plants are grown from pieces of a potato placed in the ground. This method of reproduction is most similar to
A. Sexual reproduction
B. Cloning
C. Zygote formation
D. Genetic engineering

6. The graph below shows data on the average life expectancy of humans.

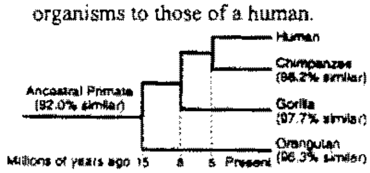


Which of the following best describes what happened to life expectancy from 1910 to 1970?
A. Life expectancy increased sharply, then leveled off
B. Life expectancy increased gradually over time
C. Life expectancy decreased gradually over time
D. Life expectancy did not change.

7. Fossils of an **extinct** species of giant armadillo were found to be similar to a smaller species of armadillo **presently** inhabiting the same region. This similarity could best be explained on the basis of

A. Evolution from older forms
B. Geographic Isolation
C. The Joyce theory
D. One species going extinct and a new species appearing in its place

8. The diagram shows a comparison of nitrogen base sequences in the DNA of some organisms to those of a human.



According to this diagram, humans are most closely related to the

A. ancestral primate
B. chimpanzee
C. gorilla
D. orangutan

9. A classification system is shown below

Classification	Examples
kingdom-Animal	Dolphin, house cat, song-bird, lynx wolf, earthworm, butterfly, hydra
Phylum-Chordata	Dolphin, house cat, songbird, lynx wolf
Genus- Felis	house cat, lynx
Species- domestica	house cat

This classification scheme shows that the house cat is most related to the

A. Earthworm
B. Dolphin
C. Song-bird
D. Hydra

10. Which change would usually ***increase*** competition for food among the squirrel population in a certain area?

A. An epidemic of rabies among squirrels
B. An increase in the number of squirrels killed on highways
C. An increase in the number of hawks that prey on squirrels
D. A temporary increase in the reproductive rate of squirrels

11. Darwin's theories can best be summed up by the following phrase:
A. Organisms survive because they are the fastest
B. Organisms that are best suited to their environment survive to reproduce and pass on their traits
C. Survival of organisms is random and does not depend on the environment
D. Survival of the slowest

12. In a certain variety of chickens, the gene for black feather color and genes for white feather color are **co-dominant**. This variety of chicken will most likely have
A. Three possible phenotypes for color
B. White feather color only
C. Only two genotypes for feather color
D. Black feather color only

13. Kernel color in corn is a trait determined by two alleles. The Dominant allele (P) produces a purple color, and the recessive allele (p) produces a yellow color. The diagram below shows an ear of corn produced by crossing two corn plants. The shaded kernels are purple, and the unshaded ones are yellow.

The yellow kernels can best be described as
A. Homozygous dominant
B. Heterozygous
C. Mixed
D. Homozygous recessive

Free Response Section: Look at the information given and answer the questions that relate.

14. One of the birds found on the Galapagos Islands is the "medium ground finch". These birds prefer to eat small seeds, which are easier to eat than large seeds. However, when food is scarce, such as during a drought, **some** of the birds can eat larger seeds. The ability to eat larger seeds is related to beak thickness, an inherited characteristic. Birds with thicker beaks are able to crush seeds more easily.

A. Describe the changes that would occur in the medium ground finch population during a long period of drought when food is scarce.

B. Explain how this set of changes is an example of the process of NATURAL SELECTION.

15. Three biology students wanted to find out if adding fertilizer to some potting soil would affect the germination of radish seeds. Each student added an equal amount of potting soil from the same bag to each of 10 cups. Student A added 1 gram of fertilizer to each cup of soil in group A. Student B added 2 grams of fertilizer to each cup of soil in group B. Student C added 3 grams of fertilizer to each cup of soil in group C. After stirring the mixture to obtain an even distribution of fertilizer, 8 radish seeds were placed in each cup and covered with 0.5 centimeter of soil. Over the next 6 days, all conditions, including the amounts of water and sunlight, were kept the same. The results are recorded in the data table below.

Data Table

	Total Number of Seedlings Visible Above the Soil		
Days After Planting	Group A	Group B	Group C
1	0	0	0
2	5	7	0
3	10	14	0
4	17	24	0
5	20	40	0
6	30	52	0

A. You must construct a line graph for the GROUP A data on the grid below. Put the "Days after planting" on the X-axis and "Number of Seedlings…" on the Y-axis.

B. According to the Data in the Table, which amount of fertilizer produced the greatest amount of seedlings?

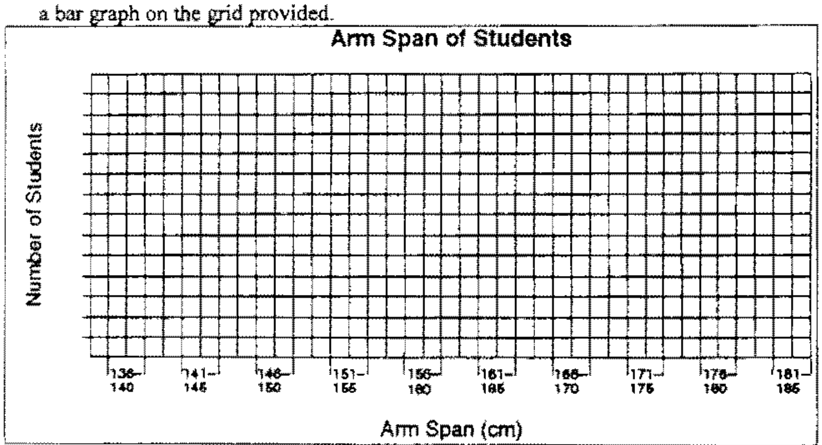
C. Form a hypothesis that explains why GROUP C did not produce any seedlings.

16. A science class was studying various human physical characteristics in an investigation for a report on human genetics. As part of the investigation, the students measured the arm span of the class members. The data table below summarizes the class results.

Arm Span of the Students

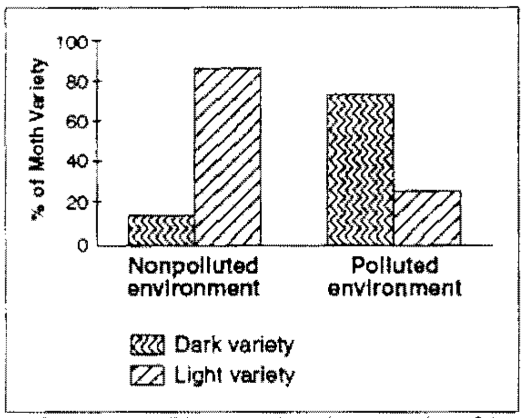
Student Arm Span (cm)	Number of Students
136–140	1
141–145	2
146–150	0
151–155	4
156–160	5
161–165	8
166–170	5
171–175	5
176–180	3
181–185	1

A. Using the information in the data table, construct a bar graph on the grid provided.



B. What conclusion can you draw from the Data and Graph?

17. Color in peppered moths is controlled by genes. A light-colored variety and a dark-colored variety of a peppered moth species exist in nature. The moths often rest on tree trunks, and several different species of birds are predators of this moth. Before industrialization in England, the light-colored variety was much more abundant than the dark-colored variety and evidence indicates that many tree trunks at that time were light in color.
Later, industrialization developed and brought pollution leaving the tree trunks covered with dark-colored soot. The results of a study made in England are shown below.



A. State one possible reason that a larger number of the dark-colored variety were present in the polluted environment.

B. State one possible reason that the light-colored variety was not completely eliminated from the polluted environment.

C. During the past few decades, air pollution control laws in many areas of England greatly limited the soot and other air pollutants coming from the burning of coal. State one way the decrease in soot and other air pollutants will most likely influence the survival of the light-colored variety of peppered moth.

D. Which conclusion can best be drawn from the information given?

(1) The trait for dark coloration better suits the peppered moth for survival in non-polluted environments.
(2) The trait for light coloration better suits the peppered moth for survival in polluted environments.
(3) The variation of color in the peppered moth has no influence on survival of the moth.
(4) A given trait may be a favorable adaptation in one environment, but not in another environment.

18. In an investigation, plants of the same species and the same initial height were exposed to a constant number of hours of light each day. The number of hours per day was different for each plant, but all other environmental factors were the same. At the conclusion of the investigation, the final height of each plant was measured. The following data were recorded:

8 hours, 25 cm; 4 hours, 12 cm; 2 hours, 5 cm; 14 hours, 35 cm;
12 hours, 35 cm; 10 hours, 34 cm; 6 hours, 18 cm

A. Organize the data by completing both columns in the data table provided, so that the hours of daily light exposure increase from the top to the bottom of the table.

Data Table

Daily Light Exposure (hours)	Final Height (cm)

B. Form a hypothesis that can explain why the plant exposed to light 24 hours a day was the shortest.

C. If another plant of the same species had been used in the investigation and exposed to 16 hours of light per day, what would the final height of the plant probably have been? Support your answer.

Appendix 5: Sample Exam With Low Proportion Of High Cognitive Demand Items

[redacted] BIOLOGY (Course [redacted]; Section [redacted]
FINAL EXAM
1st Semester

Directions: Select the best answer to the following questions. DO NOT WRITE ON THIS EXAM!

Questions:

1. Which of the following parts of your *Modern Biology* textbook would be most useful for finding the definition of the term organelle? (SG 11, CAS A)

a. table of contents
b. glossary
c. index
d. Appendix: Measurement

2. Using only your sense of sight, identify the following statement as an observation, inference, both, or neither. M & Ms have a chocolate-flavored inside. (SG 11, CAS B)

a. observation
b. inference
c. both a and b
d. neither a nor b

3. Which of the following correctly pairs a theme of biology with its description? (SG 12, CAS A, CAS B, CAS C)

a. Stability and homeostasis—living things maintain very stable internal conditions.
b. Reproduction and inheritance—all organisms are composed of cells which have special internal structures that carry out life processes.
c. Evolution—organisms are dependent on one another and their environment.
d. Cell structure and function—organisms transmit hereditary information (in the form of DNA) to their offspring.

4. The scientific process that involves using the five senses is (SG 11, CAS B)

a. inference.
b. analyzing.
c. modeling.
d. observation.

5. The most important driving force in evolution is (SG 12, CAS A)

a. natural selection.
b. heterotrophy.

c. autotrophy.
d. asexual reproduction.

6. The organism shown at the right is a (SG 12, CAS A)

a. unicellular autotroph.
b. unicellular heterotroph.
c. multicellular autotroph.
d. multicellular heterotroph.

7. The element nitrogen has: (SG 12, CAS C)

a. 5 valence electrons.
b. 6 total protons.
c. 4 total electrons.
d. 5 total neutrons.

8. The number of neutrons present in an atom can be calculated by: (SG 12, CAS C)

a. adding the number of protons and electrons.
b. subtracting the number of electrons from the number of protons.
c. subtracting the number of protons (atomic number) from the atomic mass.
d. dividing the atomic mass by the number of protons (atomic number).

9. The nucleus of an atom contains (SG 12, CAS C)

a. neutrons ONLY.
b. electrons ONLY.
c. electrons and protons.
d. protons and neutrons.

10. Valence electrons (SG 12, CAS C)

a. are the electrons present on the innermost electron shell of an atom.
b. are gained, lost, or shared between different elements to form compounds.
c. are located within the nucleus of the atom.
d. can be added to the number of neutrons to calculate atomic mass of an atom.

11. In covalent bonds, electrons are ______________________. In ionic bonds, electrons are ________________________. (SG 12, CAS C)

a. transferred; destroyed
b. shared; transferred
c. shared; duplicated
d. transferred; shared

12. Solution X has a pH of 5.40. Solution X (SG 12, CAS C)

a. is a base.
b. is a neutral solution.
c. is an acid.
d. has a similar pH to pure water.

13. Negatively charged particles that move around an atom's nucleus are (SG 12, CAS C)

a. electrons.
b. protons.
c. ions.
d. neutrons.

14. An amino acid is (SG 12, CAS A)

a. an attractive force between like particles.
b. an attractive force between unlike particles.
c. a component of many lipids.
d. a monomer of proteins.

15. A bond that forms between a positively charged hydrogen atom of one molecule and a negatively charged region of another molecule is a(n) (SG 12, CAS C)

a. ionic bond.
b. hydrogen bond.
c. covalent bond.
d. basic bond.

16. The presence of four electrons in the outermost energy level of a carbon atom enables (SG 12, CAS C)

a. carbon atoms to form four covalent bonds with atoms of other elements.
b. carbon atoms to form covalent bonds with other carbon atoms.
c. carbon atoms to form double bonds with other atoms.
d. All of the above.

17. This class of organic compounds has a monosaccharide monomer and functions in providing organisms with a quick source of energy. (SG 12, CAS A)

a. Nucleic acids
b. Lipids

c. Carbohydrates
d. Amino acids

18. Which of the following statements is not a tenet of the Cell Theory? (SG 12, CAS A)

a. All living things are composed of one or more cells.
b. New cells come from reproduction of existing cells.
c. The daily activities necessary for life occur within cells.
d. All types of cells have a special membrane-bound organelle called a nucleus.

19. The cell membrane is selectively permeable. What does this mean? (SG 12, CAS A)

a. The cell nucleus stores hereditary information in the form of DNA.
b. Cells have golgi bodies which process and package substances made by the cell.
c. The cell membrane is made of a lipid bilayer.
d. The cell membrane controls the ease and extent to which substances enter and exit the cell.

20. Cells that have a high surface area to volume ratio (SG 12, CAS A)

a. have a difficult time obtaining nutrients and excreting waste.
b. have an easier time obtaining nutrients and excreting waste than cells with a low surface area to volume ratio.
c. are rarely found in living things.
d. All of the above.

21. In which of the following types of cells would you expect to find a large number of mitochondria? (SG 12, CAS A)

a. bone
b. skin
c. muscle
d. blood

Use the diagram of a cell below to answer questions 22-24.

22. Identify the ribosome. (SG 12, CAS A)

23. Identify the golgi bodies (golgi apparatus). (SG 12, CAS A)

24. Identify the cytoplasm. (SG 12, CAS A)

25. A type of transport in which water moves across a membrane and down its concentration gradient is (SG 12, CAS A)

a. simple diffusion.
b. facilitated diffusion.
c. diffusion through ion channels.
d. osmosis.

26. The solution inside a cell is 25% salt and 75% water. This cell is placed in a beaker which has a 25% salt and 75% water solution. What will happen in this scenario? (SG 12, CAS A)

a. There will be no net movement of water in this situation.
b. Water will move out of the cell into the solution in the beaker.
c. Water will move from the solution in the beaker into the cell.
d. The cell would lose so much water that it would undergo plasmolysis.

27. Which of the following are examples of passive transport processes (which do not require energy)? (SG 12, CAS A)

a. endocytosis, diffusion
b. exocytosis, endocytosis
c. diffusion, osmosis
d. sodium-potassium pump, osmosis

28. A solution that has a relatively high solute (salt) concentration and a relatively low solvent (water) concentration can be described as: (SG 12, CAS A)

a. hypotonic.
b. hypertonic.
c. isotonic.
d. neutral.

29. The following diagram represents the process of (SG 12, CAS A)

a. endocytosis.
b. exocytosis.
c. the action of the sodium-potassium pump.
d. cellular reproduction.

Use the diagram of the chloroplast to answer questions 30-31.

30. Letter B signifies the (SG 12, CAS A)

a. thylakoid.
b. granum.
c. stroma.
d. cytoplasm.

31. Letter D signifies the (SG 12, CAS A)

a. thylakoid.
b. granum
c. stroma
d. cytoplasm.

32. Photosynthesis is an example of a biochemical pathway. What is a biochemical pathway? (SG 12, CAS B)

a. a three-carbon molecule in the Calvin cycle
b. the component colors of white light
c. a series of chemical reactions in which the product of one reaction is used in the next
d. a cluster of pigment molecules located in the thylakoid membrane

33. The anaerobic pathways provide enough energy to meet all of the energy needs of (SG 12, CAS B)

a. all organisms.
b. no unicellular and most multicellular organisms.
c. most unicellular and some multicellular organisms.
d. no organisms.

34. Photosystem I and II, the primary electron acceptors, and the electron transport chain are all located in the (SG 12, CAS B)

a. thylakoid membrane.
b. stroma.

c. outer membrane of the chloroplast.
d. chlorophyll.

35. The electrons necessary for Photosystem II to operate come from (SG 12, CAS B)

a. photosystem I.
b. enzymes present in the chlorophyll.
c. carbohydrates.
d. the splitting of a water molecule.

36. The starting substance used in the first step of the Krebs cycle, which is regenerated in the last step of the cycle, is (SG 12, CAS B)

a. acetyl CoA.
b. pyruvic acid.
c. oxaloacetic acid.
d. citric acid.

37. The following environmental condition determines whether fermentation or the Krebs cycle occurs following glycolysis: (SG 12, CAS B)

a. If oxygen is present after glycolysis, the Krebs cycle will occur.
b. If oxygen is present after glycolysis, fermentation will occur.
c. If no oxygen is present after glycolysis, the Krebs cycle will occur.
d. None of the above are true statements.

For questions 38-40, identify the statements as true or false.

38. The process of glycolysis produces a lot of ATP, whereas the Electron Transport Chain (ETC) produces lesser amounts of ATP. (SG 12, CAS B)

a. true
b. false

39. Oxygen is important to the functioning of the ETC during cellular respiration because it acts as the final electron acceptor and allows electrons to keep moving along the chain. (SG 12, CAS B)

a. true

b. false

40. Fermentation is such an important process because it results in the production of large amounts of ATP. (SG 12, CAS B)

a. true
b. false

41. DNA in a condensed form wrapped around histone proteins is referred to as (SG 12, CAS A)

a. a chromosome.
b. chromatin.
c. a centromere.
d. a centrosome.

42. Is the diagram below representative of a haploid or diploid cell? (SG 12, CAS A)

a. haploid
b. diploid

43. In the cell cycle, interphase is composed of (SG 12, CAS A)

a. 2 stages: M phase and cytokinesis
b. 3 stages: M phase, G_1 phase, G_2 phase
c. 3 stages: G_0 phase, G_1 phase, G_2 phase
d. 3 stages: G_1 phase, S phase, G_2 phase

44. The DNA is copied: (SG 12, CAS A)

a. during cell division.
b. during the S phase.
c. during cytokinesis.
d. during the G_1 phase.

45. The four stages of mitosis in the order that they occur are: (SG 12, CAS A)

a. prophase, metaphase, anaphase, telophase.
b. prophase, anaphase, metaphase, telophase.
c. prophase, metaphase, telophase, anaphase.
d. prophase, telophase, anaphase, metaphase.

46. In __________________ reproduction, the offspring are genetically identical to their parents. (SG 12, CAS A)

a. sexual
b. asexual

47. Which of the following events occurs during synapsis? (SG 12, CAS A)

a. replication of the DNA
b. appearance of spindle fibers
c. division of the cytoplasm
d. pairing of homologous chromosomes to form a tetrad

48. Humans have 46 chromosomes in all cells except sperm and egg cells. How many of these chromosomes are autosomes? (SG 12, CAS A)

a. 2
b. 23
c. 44
d. 46

49. Between cell divisions, the DNA in a eukaryotic cell is uncoiled and spread out; in this form it is called (SG 12, CAS A)

a. chromatid.
b. chromatin.
c. histone.
d. nonhistone.

50. In the G_0 phase, cells (SG 12, CAS A)

a. synthesize DNA.
b. prepare for cell division.
c. exit from the cell cycle.
d. move their chromosomes to the cell equator.

Cited Literature

AAAS Project 2061. (2005). High School Biology Textbooks: A Benchmarks-Based Evaluation. Retrieved February 2, 2014, from http://www.project2061.org/publications/textbook/hsbio/report/default.htm

Abell, S. K., & Lederman, N. G. (Eds.). (2007). *Handbook Of Research On Science Education*. Mahwah, NJ: Lawrence Erlbaum Associates.

ACT Inc. (2007). Rigor At Risk: Reaffirming Quality In The High School Core Curriculum (pp. 60). Iowa City, IA: ACT.

ACT Inc., & The Education Trust. (2004). On Course For Success: A Close Look At Selected High School Courses That Prepare Students For College And Work (pp. 83). Washington, DC.

Adelman, C. (1999). Answers in the Tool Box: Academic Intensity, Attendance Patterns, and Bachelor's Degree Attainment. Washington, DC: U. S. Department of Education.

Allensworth, E. (2005). Graduation and Dropout Trends in Chicago: A Look at Cohorts of Students from 1991 to 2004. Chicago, IL: Consortium on Chicago School Research.

Allensworth, E., Correa, M., & Ponisciak, S. (2008a). ACT Preparation: Too Much, Too Late. Chicago, IL: Consortium on Chicago School Research.

Allensworth, E., Correa, M., & Ponisciak, S. (2008b). From High School To The Future: ACT Preparation: Too Much, Too Late. Chicago: Chicago Consortium On School Research.

Allensworth, E., & Easton, J. Q. (2005). The On-Track Indicator as a Predictor of High School Graduation. Chicago, IL: Consortium On Chicago School Research at the University Of Chicago.

Allensworth, E. M., & Easton, J. Q. (2007). What Matters for Staying On-Track and Graduating in Chicago Public High Schools. Chicago, IL: Consortium on Chicago School Research.

American Association for the Advancement of Science. (1994). *Benchmarks For Science Literacy*. Washington, DC: Oxford University Press.

Arons, A. (1990). *A Guide To Introductory Physics Teaching*. New York: John Wiley and Sons.

Atkin, J. M., Black, P., & Coffey, J. (Eds.). (2001). *Classroom Assessment and the National Science Education Standards*. Washington, DC: The National Academy Press.

Ball, D. L., & Cohen, D. K. (1996). Reform By The Book: What Is--Or Might Be--The Role Of Curriculum Materials In Teacher Learning And Instructional Reform? *Educational Researcher, 25*(9), 6-8, 14.

Banilower, E. R., Boyd, S. E., Pasley, J. D., & Weiss, I. R. (2006). Lessons From A Decade Of Mathematics And Science Reform: A Capstone Report for the Local Systemic Change through Teacher Enhancement Initiative (pp. 1-88). Chapel Hill, NC: Horizon Research, Inc.

Berwick, D. M. (1996). A Primer On Leading The Improvement Of Systems. *British Medical Journal (BMJ), 312 (7031)*, 619-622.

Biggs, A., Kaskel, A., Lundgren, L., & Mathieu, D. (1998). *Biology Living Systems*. New York, NY: Glencoe/McGraw-Hill.

Blank, R. K., Birman, B., Garet, M., Yoon, K. S., Jacobson, R., Smithson, J., & Minor, A. (2005). Longitudinal Study Of The Effects Of Professional Development On Improving Mathematics And Science Instruction (MSP PD Study) Year 2 Progress Report. Washington, DC: Council Of Chief State School Officers.

Blank, R. K., Porter, A. C., & Smithson, J. (2001). New Tools for Analyzing Teaching, Curriculum and Standards in Mathematics & Science (pp. 95): Council of Chief State School Officers.

Board On Science Education. (2014). *Implementing the Vision of the Framework and Next Generation Science Standards: A Guide for Science Educators*. National Academy Press. Washington, DC.

Bogaert, L. M. R., Fendt, C. R., & Vogele, E. (2006). Piloting Math Benchmark Assessments (pp. 20). Chicago, IL: Chicago Public Schools.

Brown, D. F. (2004). Urban Teachers' Professed Classroom Management Strategies. *Urban Education, 39*(3), 266-289. doi: 10.1177/0042085904263258

BSCS. (2006). The Cornerstone-to-Capstone Approach: Creating Coherence In High School Science (pp. 206). Colorado Springs, CO: BSCS.

Bugler, D. (2006). Prairie State Achievement Exam Over Time Report. Retrieved May 27, 2007, from http://research.cps.k12.il.us/export/sites/default/accountweb/Reports/Citywide/psae_citywide_fmt.xls

Bulkley, K. E., Christman, J. B., Goertz, M. E., & Lawrence, N. R. (2010). Building With Benchmarks: The Rolf Of The District In Philadelphia's Benchmark Assessment System. *Peabody Journal Of Education, 85*.

Bullitt County Geneological Society. (1912). 1912 Eighth Grade Examination for Bullitt County Schools. Retrieved October 3, 2013, 2013, from http://www.bullittcountyhistory.com/bchistory/schoolexam1912.html

Burch, P., & Spillane, J. P. (2005). How Subjects Matter In District Office Practice: Instructionally Relevant Policy In Urban School District Redesign. *Journal Of Educational Change, 6*, 51-76.

Bybee, R. W. (1997). *Achieving Scientific Literacy: From Purposes To Practices*: Heinemann.

C-STEMEC. (2013). Putting It All Together: Supporting K-12 STEM Education In Illinois. Chicago, IL: The University Of Chicago.

Chicago Public Schools. (2000a). *Expecting More: Program Of Study for Biology*. Chicago, IL: Chicago Public Schools.

Chicago Public Schools. (2000b). *Expecting More: Program Of Study for Chemistry*. Chicago, IL: Chicago Public Schools.

Chicago Public Schools. (2000c). *Expecting More: Program Of Study for Physics*. Chicago, IL: Chicago Public Schools,.

Chicago Public Schools. (2002). Every Child, Every School: An Education Plan for the Chicago Public Schools. Chicago, IL: Chicago Public Schools.

Amend Board Report 01-0124-PO2 Policy On High School Promotion (2004).

Chicago Public Schools. (2013). The Next Generation: Chicago's Children. Chicago, IL: Chicago Public Schools.

City, E. A., Elmore, R. F., Fiarman, S. E., & Teitel, L. (2009). *Instructional Rounds in Education: A Network Approach to Improving Teaching and Learning*. Cambridge, MA: Harvard Education Press.

Clune, W. H., & White, P. A. (2008). Policy Effectiveness of Interim Assessments in Providence Public Schools. Madison, WI: Wisconsin Center for Education Research.

Coburn, C. E., Touré, J., & Yamashita, M. (2009). Evidence, Interpretation, and Persuasion: Instructional Decision Making at the District Central Office. *Teachers College Record, 111*(4), 1115-1161.

Cohen, D. K., & Ball, D. L. (2007). Educational Innovation and the Problem of Scale. In B. Schneider & S.-K. McDonald (Eds.), *Scale Up In Education Volume I: Ideas In Principle* (Vol. 1, pp. 19-56). Lanham, MD: Rowman & Littlefield.

Committee on Assessment In Support of Instruction and Learning. (2003, January 23-24, 2003). *Assessment in Support of Instruction and Learning: Bridging the Gap Between Large-Scale and Classroom Assessment.* Paper presented at the Bridging the Gap Between Large-Scale and Classroom Assessment, Washington, DC.

Committee on Developing Assessments of Science Proficiency in K-12. (2013). *Developing Assessments For Next Generation Science Standards* (J. W. Pellegrino, M. R. Wilson, J. A. Koenig & A. S. Beatty Eds.). Washington, DC: National Academy Press.

Committee on Developing The Capacity To Select Effective Instructional Materials. (1999). *Selecting Instructional Materials: A Guide For K-12 Science* (M. Singer & J. Tuomi Eds.): National Research Council.

Committee on High School Science Laboratories: Role and Vision. (2005). *America's Lab Report: Investigations in High School Science* (S. R. Singer, M. L. Hilton & H. A. Schweingruber Eds.): National Academy Press.

Committee on Science Learning Kindergarten through Eighth Grade, Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). Taking Science To School: Learning and Teaching Science in Grades K-8 (pp. 404). Washington, DC: The National Academies Press.

Committee on Test Design for K-12 Science Achievement. (2005). *Systems for State Science Assessment* (M. R. Wilson & M. W. Bertenthal Eds.). Washington, DC: National Academies Press.

Committee on the Status Contributions and Future Directions of Discipline-Based Education Research. (2012). *Discipline-Based Education Research: Understanding and Improving Learning in Undergraduate Science and Engineering* (S. R. Singer, N. R. NIelsen & H. A. Schweingruber Eds.). National Research Council.

Community for Advancing Discovery Research in Education (CADRE) at the Education Development Center. (2012). Preparing and Supporting STEM Educators. Retrieved February 1, 2014, from http://successfulstemeducation.org/sites/successfulstemeducation.org/files/Preparing Supporting STEM Educators_FINAL.pdf

Cuban, L. (1993). *How Teachers Taught: Consistancy And Change In American Classrooms 1980-1990*. New York, NY: Teachers College Press.

Cuban, L. (2007). Hugging The Middle: Teaching In An Era Of Testing And Accountability, 1908-2005. *Education Policy Analysis Archives, 15*(1).

David, L. (2006). Science a la Joe Camel. *The Washington Post*.

DeBoer, G. (1991). *A History Of Ideas In Science Education: Implications For Practice*. New York, NY: Teachers College Press.

Deiger, M. (2003). Office of Mathematics And Science Math and Science Textbook Inventory (pp. 1-4). Chicago, IL: Chicago Public Schools Office Of Mathematics And Science.

Desimone, L. M., Porter, A. C., Garet, M. S., Yoon, K. S., & Birman, B. F. (2002). Effects of Professional Development on Teachers' Instruction: Results from a Three-year Longitudinal Study. *Educational Evaluation and Policy Analysis, 24*(2), 81-112. doi: 10.3102/01623737024002081

Dewey, J. (1916). *Democracy And Education: An Introduction To The Philosophy Of Education*: Macmillan.

Donnelly, L. A., & Sadler, T. D. (2009). High School Science Teachers' Views Of Standards And Accountabilty. *Science Education*. doi: 10.1002/sce.20347

Dougherty, C., Mellor, L., & Jian, S. (2006). Orange Juice or Orange Drink: Ensuring That "Advanced Courses" Live Up To Their Labels (pp. 10). Austin, TX: National Center for Educational Accountability.

Dow, P. (2000). Inquiry and the National Science Education Standards: A Guide For Teaching And Learning (pp. 202): National Academy of Sciences.

Doyle, W. (1983). Academic Work. *Review Of Educational Research, 53*(2), 159-199.

Dufour, R., & Eaker, R. E. (1998). *Professional Learning Communities at Work: Best Practices for Enhancing Student Achievement*: Solution Tree.

Duncan, A. (2010). Beyond the Bubble Tests: The Next Generation of Assessments. Retrieved January 9,, 2011, from http://www.ed.gov/news/speeches/beyond-bubble-tests-next-generation-assessments-secretary-arne-duncans-remarks-state-l

Edelson, D. C. (2007). Environmental Science For All? Considering Environmental Sciecne for Inclusion In the High School Core Curriculum. *Science Educator, 16*(1), 42-56.

Elmore, R. F. (1999). Building A New Structure For School Leadership. *American Educator*.

Elmore, R. F. (2000). Building A New Structure For School Leadership (pp. 42): The Albert Shanker Insitiute.

Elmore, R. F. (2002). Bridging The Gap Between Standards And Achievement: The Imperative For Professional Development In Education (pp. 44): The Albert Shanker Institute.

Elmore, R. F. (2004). *School Reform From The Inside Out: Policy, Practice, And Performance*. Cambridge, MA: Harvard Education Press.

Farmer, C., & Selch, E. (2010, January 14). Are final exams necessary at the end of the semester? *Chicago Tribune*. Retrieved from http://articles.chicagotribune.com/2010-01-14/news/1001150311_1_final-exams-student-s-grade-tests

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing Organizational Routines as a Source of Flexibility and Change. *Administrative Science Quarterly, 48*(1), 94-118.

Feranchak, B. (2003). CMSI Evaluation Framework (pp. 20). Chicago, IL: Chicago Public Schools Office of Mathematics And Science.

Feranchak, B., & Price, K. (2004). CMSI Evaluation Framework (pp. 24). Chicago, IL: Chicago Public Schools Evaluation Framework.

Feranchak, B., Price, K., & Lemke, M. (2005). CMSI Evaluation Framework (pp. 29). Chicago, IL: Chicago Public Schools Office Of Mathematics And Science.

Ferkenhoff, E. (2006, May 12, 2006). The Greening Of Chicago. *Time,* 1.

Finn, C. E., Petrilli, M. J., & Julian, L. (2006). The State Of State Standards 2006 (pp. 121). Washington, DC: Fordham Foundation.

Gamoran, A. (2004, June 3, 2004). *Organizational Conditions that Support Inquiry In High School Science Instruction.* Paper presented at the Second Fact-Finding Meeting of the National Academy of Science Committee On High School Science Laboratories: Role and Vision, Washington, DC.

Gamoran, A., Anderson, C. W., Quiroz, P. A., Secada, W. G., Williams, T., & Ashmann, S. (2003). *Transforming Teaching In Math And Science: How Schools And Districts Can Support Change*. New York, NY: Teachers College Press.

Gamoran, A., Porter, A. C., Smithson, J., & White, P. A. (1997). Upgrading High School Mathematics Instruction: Improving Learning Opportunities for Low-Achieving, Low-Income Youth. *Educational Evaluation and Policy Analysis, 19*(4), 325. doi: 10.3102/01623737019004325

Goertz, M. E., Oláh, L. N., & Rigan, M. (2009). From Testing to Teaching: The Use of Interim Assessments in Classroom Instruction (Vol. CPRE Research Report # RR-65, pp. 251): Consortium for Policy Research in Education.

Gonzales, P., Guzmán, J. C., Partelow, L., Pahlke, E., Jocelyn, L., Kastberg, D., & Williams, T. (2004). Highlights From the Trends in International Mathematics and Science Study (TIMSS) 2003. Washington, DC: U. S. Department Of Education Institute Of Education Sciences.

Goren, P. (2010). Interim Assessments As A Strategy For Improvement: Easier Said Than Done. *Peabody Journal Of Education, 85*, 125-129. doi: 10.1080/01619561003673938

Graeber, A. O., Newton, K. J., & Chambliss, M. J. (2012). Crossing the Borders Again: Challenges in Comparing Quality Instruction in Mathematics and Reading. *Teachers College Record, 114*, 30.

Grayson, D. J. (1996). Concept Substitutuion: A Strategy For Promoting Conceptual Change. In D. F. Treagust, R. Duit & B. K. Fraser (Eds.), *Improving Teaching and Learning In Science and Mathematics* (pp. 152-161). New York, NY: Teachers College Press.

Gross, P. R., Goodenough, U., Lerner, L. S., Haack, S., Schwartz, M., Schwartz, R., & Finn, C. E. (2005). The State Of State Science Standards (pp. 75). Washington, DC: The Thomas B. Fordham Foundation.

Hallman, S. R. (2004). Chicago Teachers Project: Everyday Math Leadership Training, January to June, 2004 (pp. 13). Chicago, IL: University of Illinois at Chicago.

Hallman, S. R., Fendt, C. R., & Wenzel, S. A. (2003). Interim Evaluation Report: CMSI/CUSP Leadership Academy (pp. 35). Chicago, IL: University Of Illinois At Chicago.

Halverson, R., Feinstein, N. R., & Neshoulam, D. (2011). School Leadership For Science Education. In G. E. DeBoer (Ed.), *The Role of Public Policy In K-12 Science Education* (pp. 397-431). Charlotte, NC: Information Age Publishing.

Halverson, R. R. (2003). Systems of Practice: How Leaders Use Artifacts to Create Professional Community in Schools. *Education Policy Analysis Archives, 11*(37), 50.

Hammer, D. (1997). Discovery Learning and Discovery Teaching. *Cognition and Instruction, 15*(4), 485-529.

Hart, P. (2007). Environmental Education. In S. K. Abell & N. Lederman (Eds.), *Handbook of Research on Science Education* (pp. 689-726). Mahwah, NJ: Lawrence Erlbaum Associates.

Hatch, J. A. (2002). *Doing Qualitative Research In Educational Settings*. Albany, NY: State University of New York Press.

Heiton, L. (2014). Illinois State Board Votes in Favor of Common Science Standards. Retrieved February 9, 2014, from http://blogs.edweek.org/edweek/curriculum/2014/01/illinois_poised_to_adopt_commo.html

Hill, P. (2013). Defining and Organizing For School Autonomy. Center for Reinventing Public Education: The University of Washington.

Illinois State Board Of Education. (1985). Illinois Learning Standards - Science. Retrieved May 25, 2007, from http://www.isbe.net/ils/science/standards.htm

Illinois State Board Of Education. (1997). Illinois Learning Standards - Science. Retrieved August 9, 2013, from http://www.isbe.net/ils/science/standards.htm

Illinois State Board Of Education. (2001). Illinois Learning Standards - Science: Classroom Assessments and Performance Descriptors. Retrieved May 25, 2007, from http://www.isbe.net/ils/science/capd.htm

Illinois State Board Of Education. (2005). Illinois Science Assessment Framework PSAE Grade 11 (pp. 19). Springfield, IL: Illinois State Board of Education.

Illinois State Board Of Education. (2006). Illinois Science Assessment Framework PSAE Grade 11 (pp. 19). Springfield, IL: Illinois State Board of Education.

Illinois State Board Of Education, & ACT. (2007). Prairie State Achievement Examination Technical Manual (pp. 113). Springfield, IL: Illinois State Board of Education.

Jackson, W., Wenzel, S., Shortino, C., Tregerman, H., Casanovas, J., Cram, K., . . . Sims, A. (2013). Implementing the Next Generation Science Standards: Hallmarks of a Fully Realized School System. Chicago, IL: C-STEMEC.

Kahle, J. B. (2007). Systemic Reform: Research, Vision, and Politics. In N. G. Lederman & S. K. Abell (Eds.), *Handbook of Research on Science Education* (pp. 911-941). Mahwah, NJ: Lawrence Erlbaum Associates.

Kahne, J. E., Sporte, S. E., de La Torre, M., & Easton, J. Q. (2006). Small High Schools on a Larger Scale: The First Three Years of the Chicago High School Redesign Initiative (pp. 68 pages). Chicago, IL: Consortium on Chicago School Research.

Kanold, T. (2003, February 4, 2003). [Summer Leadership Academy for Department Chairs And Critical CPS School Leaders In Mathematics And Science].

Kucey, S., & Parsons, J. (2012). Linking Past And Present: John Dewey And Assessment For Learning. *Journal Of Teaching And Learning, 8*(1), 107-116.

Lach, M. (2004, August 2, 2004). [2005 ISAT and PSAE Science Information].

Lach, M. (2006). *K-8 Mathematics Performance Update*. Chicago, IL: Chicago Public Schools.

LaForce, M. (2004). Characteristics of Chicago Public Schools' High School Math and Science Teachers In The 2002-03 School Year (pp. 1-17). Chicago, IL: Chicago Public Schools Office Of Mathematics And Science.

LaForce, M., & Feranchak, B. (2003). Preliminary Analyses of course taking patterns of CPS high school students in math and science from 2000-2002 (pp. 8). Chicago, IL: Chicago Public Schools.

Lederman, N. (1998). The State Of Science Education: Subject Matter Without Context. *Electronic Journal Of Science Education, 3*(2).

Lederman, N. G. (1998). The State of Science Education: Subject Matter Without Context. *Electronic Journal Of Science Education, 3*(2).

Lee, V. E., Smith, J. B., Perry, T. E., & Smilie, M. A. (1999). Social Support, Academic Press, and Student Achievement: A View From The Middle Grades In Chicago. Chicago, IL: Consortium On Chicago School Research.

Luppescu, S., Allensworth, E. M., Moore, P., Torre, M. d. l., Murphy, J., & Jagesiv, S. (2011). Trends In Chicago's Schools Accross Three Eras Of Reform. Chicago, IL: Consortium on Chicago School Research at the University of Chicago.

Luppescu, S., & Hart, H. (2005). Improving Chicago's Schools: Sample High School Details 2005 (pp. 72). Chicago, IL: Consortium on Chicago School Research.

Lyons, K. C. (2013). 2012 National Survey of Science and Mathematics Education: Status Of High School Biology (pp. 29). Chapel Hill, NC: Horizon Research, Inc.

Martínez, J. F., Borko, H., & Stecher, B. M. (2011). Measuring Instructional Practice in Science Using Classroom Artifacts: Lessons Learned From Two Validation Studies. *Journal Of Research In Science Teaching, 49*(1), 38-67.

Mather, A. (2014, January 23). [Re: Brain Picking].

Max, J., & Glazerman, S. (2014). Do Disadvantaged Students Get Less Effective Teaching? Key Findings From Recent Institute Of Education Sciences Studies. Washington, DC: National Center for Education Evaluation and Regional Assistance.

Mazzeo, C. (2010). College Prep For All? What We've Learned From Chicago's Efforts. Chicago: Chicago Consortium On School Research.

McKinsey & Company. (2001). *CPS Human Capital Initiative Steering Committee*. Presentation. McKinsey & Co. Chicago, IL.

Members of the 2005 "Rising Above the Gathering Storm" Committee. (2010). Rising Above the Gathering Storm, Revisited: Rapidly Approaching Category 5. Washington, DC: National Academy of Sciences, National Academy of Engineering, and Institute of Medicine.

Miles, K. H., Hornbeck, M., & Fermanich, M. (2002). Chicago Public Schools Professional Development Project (pp. 20). Chicago, IL: Chicago Public Schools.

Miller, K., & Levine, J. (1998). *Biology*. Englewood Cliffs, NY: Prentice Hall.

Minner, D. D., Levy, A. J., & Century, J. (2010). Inquiry-Based Science Instruction—What Is It and Does It Matter? Results from a Research Synthesis Years 1984 to 2002. *Journal Of Research In Science Teaching, 47*(4), 474-496.

Monastersky, R. (2005). Why Many Biology Teachers Aren't Teaching Evolution. *The Chronicle Of Higher Education, 51*(46), 1.

Monk, D. H. (1994). Subject Area Preparations of Secondary Mathematics And Science Teachers And Student Achievement. *Economics Of Education Review, 13*(2), 125-145.

Montgomery, N., Allensworth, E., & Correa, M. (2010). Passing Through Science: The Effects Of Raising Graduation Requirements In Science On Course-Taking and Academic Achievement In Chicago. Chicago: Chicago Consortium on School Research.

National Center for Education Statistics. (2011). Science 2009: Trial Urban District Assessment Results At Grades 4 and 8. Washington, DC: Institute for Education Sciences, The U. S. Department of Education.

National Research Council. (1996). *National Science Education Standards*. Washington, DC: National Academies Press.

National Research Council. (2000). *Inquiry and the National Science Education Standards: A Guide For Teaching And Learning*. Washington, DC: The National Academies Press.

National Research Council. (2012). A Framework For K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas. In H. Quinn, H. A. Schweingruber & T. Keller (Eds.), (pp. 400). Washintgon, DC: National Academy Press.

National Science Foundation. (2006). NSF Awards $1.8 Million to Study High-School Advanced Placement Work in Math and Science [Press release]. Retrieved from http://www.nsf.gov/news/news_summ.jsp?cntn_id=106929

National Science Teachers Association. (1996). *Scope, Sequence, and Coordination: A Framework for High School Science Education*. Arlington, VA: National Science Teachers Association.

NGSS Lead States. (2013a). The Next Generation Science Standards. Retrieved September 28, 2013, 2013, from http://www.nextgenscience.org

NGSS Lead States. (2013b). *Next Generation Science Standards: For States, By States*. Washington, DC: National Academies Press.

Northern Illinois University. (2006). Keeping Illinois Competitive (pp. 127). Dekalb, IL: Northern Illinois University.

Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). Knowing What Students Know: The Science and Design Of Educational Assessment (pp. 366). Washington, DC: National Academy of Science.

Perie, M., Marion, S., & Gong, B. (2007). A Framework For Considering Interim Assessments (pp. 43). Dover, NH: The National Center For The Improvement Of Educational Assessment.

Ponisciak, S. M. (2005). Understanding the Prairie State Achievement Exam: A Descriptive Report With Analysis of Student Performance (pp. 37). Chicago, IL: Consortium on Chicago School Research.

Presley, J. B., & Gong, Y. (2006). The Demographics and Academics of College Going in Illinois (pp. 48). Edwardsville, IL: Illinois Education Research Council.

Presley, J. B., White, B. R., & Gong, Y. (2005). Examining the Distribution and Impact of Teacher Quality in Illinois (pp. 24). Edwardsville, IL: Illinois Education Research Council.

Quinn, H., Schweingruber, H., Keller, T., & Krone, R. (2010). Draft Conceptual Framework For New Science Education Standards. Washington, DC: National Academy Of Sciences.

Randall, J., & George Engelhard, J. (2009). Examining Teacher Grades Using Rasch Measurement Theory. *Journal Of Educational Measurement, 46*(1), 1-18.

RMC Research Corporation. (2005). Illinois State Board of Education Mathematics and Science Partnership Program Evaluation Surveys of Enacted Curriculum Report. Denver, CO: Illinois State Board of Education.

Roderick, M., Nagaoka, J., Allensworth, E., Coca, V., Correa, M., & Stoker, G. (2006). From High School to the Future: A first look at Chicago Public School graduates' college enrollment, college preparation, and graduation from four-year colleges (pp. 120). Chicago: Consortium on Chicago School Research.

Roseman, J. E., Caldwell, A., Gogos, A., & Kurth, L. (2006). *Mapping a Coherent Learning Progression for the Molecular Basis of Heredity*. Paper presented at the National Association for Research in Science Teaching San Francisco, CA. http://www.project2061.org/publications/articles/papers/narst2006.htm

Salazar, R. (2013). Final exams in high school--unnecessary. Retrieved January 22, 2014, from http://www.chicagonow.com/white-rhino/2013/10/final-exams-in-high-school-unnecessary/

Schmidt, G. N. (2007). A CASE history of the "Chicago Academic Standards Exams", short version... With tests botched again, it's time to abolish the secrecy of high-stakes tests. *Substance News*.

Schmidt, W. H. (2005). The Role Of Curriculum. *American Educator*(Fall 2005).

Schmidt, W. H., Zoido, P., & Cogan, L. (2014). Schooling Matters: Opportunity To Learn In PISA 2012 *OECD Education Working Papers*. OECD Publishing.

Schools, C. P. (1997). Chicago Public Schools Design For High Schools (pp. 109). Chicago, IL: Chicago Public Schools.

Schwartz, M. S., Sadler, P. M., Sonnert, G., & Tai, R. H. (2008). Depth Versus Breadth: How Content Coverage in High School Science Courses Relates to Later Success in College Science Coursework. *Science Education*. doi: 10.1002/sce.20328

Sconiers, S., Isaacs, A., Higgins, T., McBride, J., & Kelso, K. R. (2003). The ARC Center Tri-State Student Achievement Study. Lexington, MA: The Consortium For Mathematics And Its Applications.

Shirley, D., & Hargreaves, A. (2006, October 4, 2006). Data-Driven To Distraction. *Education Week*.

Singer, S. R., Hilton, M. L., & Schweingruber, H. A. (Eds.). (2005). *America's Lab Report: Investigations In High School Science*. Washington, DC: National Academies Press.

Sizer, T. R. (1992). *Horace's Compromise: The Dilemma Of The American High School*: Mariner Books.

Smith, T. M., Desimone, L. M., Zeidner, T. L., Dunn, A. C., Chatt, M., & Rumyantseva, N. L. (2007). Inquiry-oriented Instruction in Science: Who Teaches That Way? *Educational Evaluation and Policy Analysis, 29*(3), 169-199. doi: 10.3102/0162373707306025

Spillane, J. P. (2004). *Standards Deviation: How Schools Misunderstand Education Policy*. Cambridge, MA: Harvard University Press.

Spillane, J. P. (2005). Primary School Leadership Practices: How The Subject Matters. *School Leadership And Management, 25*(4), 383-397. doi: 10.1080/13634230500197231

Spillane, J. P. (2006). *Distributed Leadership*. San Francisco, CA: Jossey-Bass.

Spillane, J. P., & Hopkins, M. (2013). Organizing for instruction in education systems and school organizations: how the subject matters. *Journal of Curriculum Studies, 45*(6), 721-747. doi: 10.1080/00220272.2013.810783

Sporte, S. E., Correa, M., Hart, H. M., & Wechsler, M. E. (2009). High School Reform in Chicago Public Schools: Instructional Development Systems (Vol. 2, pp. 59). Chicago, IL: SRI International.

Sporte, S. E., Kahne, J. E., & Correa, M. (2004). Notes From The Ground: Teachers, Principals, and Students' Perspectives On the Chicago High School Redesign Initiative, year two (pp. 35). Chicago, IL: Consortium on Chicago School Research.

Stanko, P. (2007). 2007 Science PSAE. Retrieved May 25, 2007, from http://www.isbe.net/assessment/pdfs/2007_Science_PSAE.pdf

Stern, L., & Alhlgren, A. (2002). Analysis of Students' Assessments in Middle School Curriculum Materials: Aiming Precisely at Benchmarks and Standards. *Journal Of Research In Science Teaching, 39*(9), 889-910.

Stevens, W. D., & Kahne, J. E. (2006). Professional Communities and Instructional Improvement Practices: A Study Of Small High Schools In Chicago (pp. 34). Chicago, IL: Consortium on Chicago School Research.

Stigler, J. W., & Hiebert, J. (1999). *The Teaching Gap: Best Ideas from the World's Teachers for Improving Education in the Classroom*. New York, NY: Free Press.

Stoelinga, S. R., Hart, H., & Schalliol, D. (2008). The Work of Chicago Public Schools' Principals. Chicago: Chicago Consortium On School Research.

Strauss, E., & Lioswski, M. (1998). *Biology: The Web Of Life*. Glenview, IL: Scott Forseman Addison Wesley.

Supovitz, J. (2009). Can high stakes testing leverage educational improvement? Prospects from the last decade of testing and accountability reform. *Journal Of Educational Change, 10*, 211-227. doi: 10.1007/s10833-009-9105-2

Tammy Kitzmiller versus Dover Area School District (United States District Court 2005).

Tang, X., Coffey, J. E., Elby, A., & Levin, D. M. (2009). The Scientific Method and Scientific Inquiry: Tensions in Teaching and Learning. *Science Education*. doi: 10.1002/sce.20366

The Chicago Sun Times. (2009). Grading scales at Chicago public high schools. Retrieved November 29, 2013, from http://www.suntimes.com/news/2683614-418/academy-career-community-achievement-chicago.html

The Geographical Data in Education (GEODE) Initiative. (2005). *Investigations in Environmental Science* (1st ed.). Mount Kisco, NY: It's About Time, Inc.

The National Board For Professional Teaching Standards. (2003). Adolescent and Young Adult Science Standards. Retrieved February 9, 2014, from http://www.nbpts.org/sites/default/files/documents/certificates/nbpts-certificate-aya-science-standards.pdf

The White House. (2012). Preparing A 21st Century Workforce: Science, Technology, Engineering, and Mathematics (STEM) In The 2013 Budget. Retrieved November 29, 2013, from http://www.whitehouse.gov/sites/default/files/microsites/ostp/fy2013rd_stem.pdf

Tobias, S., & Baffert, A. (2009). *Science Teaching As A Profession: What It Isn't, How It Could Be* (Vol. 1): Research Corporation for Scientific Advancement.

Tyack, D., & Cuban, L. (1995). *Tinkering Toward Utopia: A Century Of Public School Reform*. Cambridge, MA: Harvard University Press.

Usiskin, Z. (2013). *What Changes Should Be Made for the Next Edition of the Common Core Standards?* Paper presented at the Illinois Council of Teachers Of Mathematics (ICTM) Annual Meeting, Peoria, IL.

Venezia, A., Krist, M. W., & Antonio, A. L. (2003). Betraying The College Dream: How Disconnected K-12 and Postsecondary Education Systems Undermine Student Aspirations (pp. 72). Stanford, CA: Stanford University's Bridge Project.

Watanabe, R., & McGaw, B. (2003a). Learning For Tomorrow's World: First Results From PISA 2003. Paris, France: Organisation for Economic Co-operation and Development.

Watanabe, R., & McGaw, B. (2003b). Problem Solving For Tomorrow's World: First Measures of Cross-Curricular Competencies From PISA 2003: Organisation for Economic Co-operation and Development.

Weick, K. E. (1976). Educational Organizations as Loosely Coupled Systems. *Administrative Science Quarterly, 21*.

Wenzel, S. (2010). Systemic Reform of Math and Science Education in Chicago: An Interactive Web-Based Resource. Retrieved October 1, 2013, from http://www.luc.edu/scaleup/index.php

White House Office Of The Press Secretary. (2012). President Obama Announces Plans for a New, National Corps to Recognize and Reward Leading Educators in Science, Technology, Engineering, and Math [Press release]. Retrieved from http://www.whitehouse.gov/the-press-office/2012/07/17/president-obama-announces-plans-new-national-corps-recognize-and-reward-

Whitehurst, G. R. (2009). Don't Forget Curriculum. *Brown Center Letters On Education*(October 2009).

Willingham, D., & Grissmer, D. (2011, December 5, 2011). The teacher quality conundrum: If they are the problem, why are kids gaining in math?, Opinion. *New York Daily News*.

Wilson, M. R., & Bertenthal, M. B. (Eds.). (2005). *Systems For State Science Assessment*. Washington, DC: National Research Council.

Wilson, M. R., & Draney, K. (1999). *The SEPUP Assessment System*. Unpublished Technical Report. University of California, Berkeley. Berkeley, CA.

Wilson, S. (2013). Professional Development For Science Teachers. *Science, 340*(310). doi: 10.1126/science.1230725

Yager, R. E., & Stodghill, R. (1979). School Science In An Age Of Science. *Educational Leadership, 36*(6), 439-445.

Vita

Name	Michael Lach	
Education	Northeastern Illinois University, M. A. T., educational leadership	2002
	Teachers College Columbia University, M. A. T., science education	1995
	Carleton College, B. A., physics	1990
Professional History	Director of STEM Policy and Special Initiatives, Urban Education Institute at the University of Chicago	2011-present
	Special Assistant for Science, Technology, Engineering, and Mathematics Education, U. S. Department of Education	2009-2011
	Chief Officer of Teaching and Learning, Chicago Public Schools	2009
	Officer of High School Teaching and Learning, Chicago Public Schools	2007-2009
	Director of Mathematics and Science, Chicago Public Schools	2006-2007
	Director of Science, Chicago Public Schools	2002-2006
	Lead Curriculum Developer, Northwestern University	2000-2002
	Science Teacher, Lake View High School, Chicago, IL	1995-2000
	Physics Teacher, School Of The Future, New York, NY	1994-1995
	Director of Program Design, Teach For America, New York, NY	1993-1994
	Science Teacher, Alceé Fortier High School, New Orleans, LA	1990-1993

Selected Career Awards	Adler Planetarium Distinguished Science Educator	2009
	Radio Shack Top 100 Technology Teacher In America	2000
	Illinois Physics Teacher of the Year	1999
	National Board Certification, Adolescent/Young Adult Science	1998
Selected Publications	*Investigations In Environmental Science*, It's About Time/Herff Jones [lead developer]	2005
	Assessment Strategies for Laboratory Reports, *The Physics Teacher* [with Taoufik Nadji]	2003
	Everyone Needs A Mentor, *The Science Teacher* [with Douglas Goodwin]	2002
	Remediation, invited commentary for *The Science Teacher*	1999
	An Active Introduction To Evolution, *The American Biology Teacher* [with Michael Loverude]	1998
	An Inner-City Education, *Scientific American*	1992
Selected Presentations, Testimony, and Service	National Academy of Science Board on Science Education, member	2012-present
	U. S. House of Representatives Committee on Science and Technology, Subcommittee on Research and Science Education, hearing on "A Systems Approach to Improving K-12 STEM Education"	July 30, 2009
	National Academy of Engineering committee member on "Understanding and Improving K-12 Engineering Education in the United States", Division on Behavioral and Social Science and Education	2007-2009
	U. S. House of Representatives Committee on Science and Technology, Subcommittee on Research and Science Education, hearing on "Federal STEM Education Programs: Educators' Perspectives"	May 15, 2007
	National Academy of Sciences committee member on "The Role and Vision of High School Science Laboratories", Board on Science Education	2004-2005