“ABOUT AVERAGE”
A PRAGMATIC INQUIRY INTO SCHOOL PRINCIPALS’ MEANINGS FOR A
STATISTICAL CONCEPT IN INSTRUCTIONAL LEADERSHIP

A Thesis
Submitted to the Faculty of Graduate Studies and Research
In Partial Fulfillment of the Requirements
For the Degree of

Doctor of Philosophy
in
Education
University of Regina
by
Darryl Milburn Hunter
Regina, Saskatchewan

April 2014
Copyright 2014: Darryl M. Hunter
Darryl Milburn Hunter, candidate for the degree of Doctor of Philosophy in Education, has presented a thesis titled, “About Average” A Pragmatic Inquiry into School Principals’ Meanings for a Statistical Concept in Instructional Leadership, in an oral examination held on April 1, 2014. The following committee members have found the thesis acceptable in form and content, and that the candidate demonstrated satisfactory knowledge of the subject material.

External Examiner: *Dr. Jerald Paquette, Professor Emeritus, University of Western Ontario

Supervisor: Dr. William R. Dolmage, Educational Administration

Committee Member: Dr. Ronald Martin, Educational Psychology

Committee Member: **Dr. Larry Steeves, Educational Administration

Committee Member: Dr. Katherine Arbuthnott, Department of Psychology

Chair of Defense: Dr. Ronald Camp, Kenneth Levene Graduate School

*via SKYPE
**via Teleconference
Abstract

This mixed methods, sequential, exploratory study addresses the problem, “How significant are statistical representations of ‘average student achievement’ for school administrators as instructional leaders?” Both phases of the study were designed within Charles Sanders Peirce’s pragmatic theory of interpretation.

In the first, phenomenological phase, 10 Saskatchewan school principals were each interviewed three times and invited to read aloud three different student achievement reports. Principals generally held a “centre-of-balance” conception for the average, which related to perspectives deriving from their organizational position. Abductive reasoning, a proclivity to act upon “below average” student achievement, leadership through asking leading questions, an inquiry cast of mind, and other pragmatic principles were clearly apparent. No evidence was found that school administrators were constrained by normative statistics into a uniform outlook, nor into purely instrumental behaviour.

In a succeeding, overlapping phase based in the psychophysics of perception, Saskatchewan school leaders (principals and vice-principals) (n=210) were randomly assigned to one of four groups and asked to read an achievement report depicting student performance as a distribution of scores on a criterion scale. School leaders’ dispositions to be rational-analytical or intuitive-experiential were measured pre-and post-reading. A MANCOVA revealed small but significant changes in school leaders’ dispositions depending on the way the report was framed. Small but significant interactions between valence and audience on a reader’s rationality were observed. Negatively-framed test scores effected greater changes than positively-framed test scores in diminishing school leaders’ beliefs in their rationality. Principals’ and vice-principals’ dispositions did not differ.
I conclude that reading reports which depict student achievements within a normative distribution has little statistical significance in changing leadership practice. However, school principals’ interpretations demonstrate the substantial practical significance of statistics when leading change. School administrators consider average student achievement not with the inferential patterns assumed within contemporary notions of heuristic irrationality, but rather as a reasoned form of inquisitive thinking and behaviour that has been formalized and comprehensively described in North American philosophy for over 100 years.
Enter your skiff of Musement, push off into the lake of thought, and leave the breath of heaven to swell your sail. With your eyes open, awake to what is about or within you, and open conversation with yourself: for such is all meditation. It is, however, not a conversation in words alone, but is illustrated, like a lecture, with diagrams and with experiments.

Charles Sanders Peirce

Acknowledgements

Primary thanks are extended to Dr. Rod Dolmage, who contests the law of averages, and who has unceasingly removed blockages on the road of inquiry. Dr. Larry Steeves generously opened many doors, especially to the expansive literature on instructional leadership. Dr. Ron Martin steered me through the burgeoning research on (ir)rational decision-making, enthusiastically helping with design questions. Dr. Katherine Arbuthnott has enhanced this project’s prospects, both practically and theoretically.

The University of Regina’s Graduate Student Scholarships and Dean’s Scholarship Program represent only a portion of the support this campus has given me on so many occasions. A Jack and John Spencer Middleton Scholarship defrayed some travel expenses, while a Saskatchewan Innovation and Opportunity Graduate Scholarship enabled a link through the Social Sciences Research Laboratory at the University of Saskatchewan. A League of Educational Administrators, Directors and Superintendents award gave me confidence that the project was important. I acknowledge the Government of Saskatchewan’s tuition support, and am particularly appreciative for a Social Science and Humanities Research Council Doctoral Scholarship and a Killam Trust Pre-Doctoral Fellowship which underwrote my initial forays in this realm: this document illustrates Peirce’s dictum that probabilities play out in the long run.
Post Defence Acknowledgement

I am grateful to Dr. Jerry Paquette at the University of Western Ontario for underlining the necessity of critiquing our notions of causation, for highlighting distinctions between morality and ethics, and for emphasizing the virtues of remaining hesitant when interpreting statistics.
Dedication

The time and insights of the ten school principals I interviewed in depth and the efforts of another 210 Saskatchewan school principals and vice-principals, who completed an inventory of their beliefs while in the midst of other pressing issues, are most appreciated. Several senior administrators, on campus and off, have smoothed my path in several ways.

This project was conceived in stages through advice I received on several Canadian campuses. Dr. Scott Goble at the University of British Columbia “conducted” me into Peirce’s writings. As foil and friend, Dr. James Rude at the University of Alberta pointed out that economists’ notions of utility were the wrong starting point. At the University of Toronto, Prof. Diane Gerin-Lajoie persuaded me that qualitative methods could illuminate quantitative matters, even though their grammars differ. Dr. Trevor Gambell, now emeritus at the University of Saskatchewan, has always exemplified that the “Will to Learn” works in symbiosis with the “Will to Believe.” A thoughtful question about pragmatism and decision-making, posed years ago by Dr. Doug Stewart at the University of Regina, has sparked much reading. Across this same campus, Dr. Don Sharpe kindly helped with many statistical questions.

Of course, Judy has once again been my quiet hero, patiently tolerating yet another academic excursion. However, she insists my February 14 insight about averages will remain a gift to myself, not to her. She, like my parents and my children, has always waved encouragingly from shore while I tack across the lake of thought. But I dedicate this document to my grandparents who, though now departed, continue to whisper.
# Table of Contents

Abstract .............................................................................................................................................. i  
Acknowledgements .......................................................................................................................... iii  
Post Defence Acknowledgement ...................................................................................................... iv  
Dedication ........................................................................................................................................ v  
Table of Contents ............................................................................................................................ vi  
List of Tables ..................................................................................................................................... ix  
List of Figures .................................................................................................................................. x  
List of Abbreviations, Symbols, Nomenclature ................................................................................ xi  

Chapter 1: Introduction and Inception of the Study ................................................................. 1  
1.1 Background of the Inquirer ........................................................................................................ 1  
1.2 Beliefs of the Inquirer ............................................................................................................... 6  
1.3 Beginnings of the Inquiry ......................................................................................................... 11  

Chapter 2: The Interpretation of Average Student Achievement ........................................ 16  
2.1 Management and Meaning ....................................................................................................... 16  
2.2 Research Purpose and Questions ............................................................................................ 21  
2.3 Contexts for Considering the Average ..................................................................................... 23  
2.4 Statistics and Instructional Leadership .................................................................................... 27  
2.5 Data and Decision Making ...................................................................................................... 35  
2.6 Significance of the Study .......................................................................................................... 39  
2.7 Delimitation and Limitations .................................................................................................. 43  
2.8 Recapitulation and Organization of the Dissertation ............................................................ 47  

Chapter 3: Review of Literature on Averages .................................................................... 49  
3.1 Mathematical Forms of the Average ...................................................................................... 49  
3.2 Sociological Forms for a Norm .............................................................................................. 53  
3.3 Formulae for Computing an Average ...................................................................................... 58  
3.4 Cultural Formations Around an Average ............................................................................... 61  
3.5 Formats for Presentation of an Average .................................................................................. 68  
3.6 Functions of Statistics in Education ....................................................................................... 70  
3.7 Leadership Finesse in Sense-making ..................................................................................... 80  
3.8 Failures in Cognition ............................................................................................................... 82  
3.9 Framing and Interpretation ..................................................................................................... 87  
3.10 Recapitulation ........................................................................................................................ 90
Chapter 4: Pragmatic and Semiotic Ideas of Charles Sanders Peirce .................92

4.1 Charles Peirce’s Life and Location in North American Scholarship .............92
4.2 Pragmatism .................................................................................96
  4.21 Continuity and anti-Cartesianism ............................................96
  4.22 Belief and doubt ......................................................................101
  4.23 Perception and intuition ..........................................................103
  4.24 Induction and abduction ...........................................................105
  4.25 Diagrams and deductive reasoning ..........................................107
  4.26 Scientific inquiry and statistical cognition .................................110

4.3 Semiotics .....................................................................................114
  4.31 Phenomenology and psychology ..............................................114
  4.32 Semiotics and signs ..................................................................116
  4.33 Sign types and classes ..............................................................122
  4.34 Sign transformation and inference ............................................124
  4.35 Convergence and norms ..........................................................127
  4.36 Causation and consequences ...................................................130

4.4 Recapitulation ............................................................................134

Chapter 5: Methodology and Methods .................................................136

5.1 Presuppositions .........................................................................136
5.2 Permissions ................................................................................138
5.3 Qualitative Method for Phase 1 ..................................................140
  5.31 Assumptions .............................................................................140
  5.32 Research question and abductions ............................................144
  5.33 Respondents and sampling ......................................................145
  5.34 Interviews and instrumentation ..............................................148
  5.35 Coding and analyses ...............................................................154
5.4 Quantitative Method for Phase 2 ..................................................157
  5.41 Assumptions .............................................................................157
  5.42 Experiment design and decisions ..............................................159
  5.43 School division sampling and randomization ................................162
  5.44 Instrumentation .......................................................................165
  5.45 Data collection and entry ..........................................................169
  5.46 Data analysis techniques ...........................................................172
5.5 Presentation of Administrators’ Conceptions of the Average ...............174
5.6 Recapitulation ............................................................................176

Chapter 6: Ten Principals’ Conceptions of Average Student Achievement .......178

6.1 Conceptions and Calculations .....................................................178
6.2 Index and Symbol ........................................................................186
6.3 Gestures and Embodied Ideas ......................................................193
6.4 Sociological Norms and Mathematical Properties ..........................196
6.5 Attention and Simplification .......................................................208
6.6 Interpretive Processes and Inferences ..........................................210
6.7 Credibility and Trust .................................................................214
6.8 Leading Questions and Interlocution ..........................................217
6.9 Consequents and Antecedents ................................................................. 227
6.10 Causes and Effects .................................................................................. 232
6.11 Recapitulation .......................................................................................... 241

Chapter 7: Effects of Framing Average Student Achievement on School
Administrators’ Problem-Solving Styles ....................................................... 243
7.1 Valence and Audience ............................................................................. 249
7.2 Occupational Position and Disposition ................................................... 256
7.3 Demographics and a Descriptive Model ................................................. 257
7.4 Discussion and Implications ................................................................... 260
7.5 Recapitulation ......................................................................................... 272

Chapter 8: The Significance of Average Student Achievement in School
Administrative Practice .................................................................................. 274
8.1 Synopsis of the Study’s Purposes and Findings ....................................... 274
8.2 Conceptions and Perspectives on the Average ......................................... 277
8.3 Inferences and Inquiry ............................................................................. 287
8.4 Administrative Beliefs and Doubt ............................................................ 291
8.5 Leadership and Leading Questions .......................................................... 304
8.6 The Average as Quality and as Quantity .................................................. 308
8.7 Practical and Statistical Significance ........................................................ 311
8.8 Reflections and Further Research Questions ............................................ 321

References .................................................................................................... 329

Appendices

Appendix A: Permissions from University of Regina and University of
Saskatchewan .................................................................................................... 363
Appendix B: Informed Consent Forms ............................................................. 365
Appendix C: Permission for Use of Rational Experiential Inventory ............... 367
Appendix D: Table of Correspondence for Research Question #1, Conjectures by
Semi-structured Interview Question Numbers in Phase 1 ............................ 368
Appendix E: Phase 1: Semi-structured Interview Questions ............................ 369
Appendix F: Phase 1: Interview Data Displays ................................................. 372
Appendix G: Phase 2: Four Frames for Interpreting the Average ..................... 376
Appendix H: Epstein’s Dual Process Model .................................................... 377
Appendix I: Matrix of Findings Cross-tabulated with Data Source ................... 378
Appendix J: Recommendations for Policy, Program and Practice .................... 379
List of Figures

Figure 4.1. Peirce’s triadic conception of the Sign. .................................................. 117
Figure 4.2. Sign as perceived and conceived by three individuals with dissimilar backgrounds. ................................................................. 120
Figure 4.3. Taxonomy of Peirce’s Sign relationships. .............................................. 122
Figure 4.4. Evolving interpretant over time, resulting in changes to the conception of the Sign and its object. ........................................ 124
Figure 5.1. Mixed-method research plan for sequential exploratory inquiry into school administrators’ conception of the average: procedure and products. ................................................................................. 139
Figure 5.2. Research design into school administrators’ perceptions of average student achievement, pre- and post-disposition measure and four frames. ........................................................................ 162
Figure 6.1. Pammy’s depiction of average as imposition on four criterion levels (Interview 1); as normative curve and as benchmark (Interview 2); and as fluctuating longitudinal line in apposition to criterion-referenced measurement results (Interview 3). ......................................................... 188
Figure 6.2. Craig demonstrating elastic grey area which he mediates, with diagram on his desk indicating shift right in public expectations. ..........190
Figure 6.3. Louise triaging students for her Watch List with right arm as average bar, and left hand scooping below average students toward her. .......... 194
Figure 6.4. Phillip’s diagram of average student achievement as a neutral middle between good and bad performance. ................................................. 202
Figure 6.5. Peter demonstrating mean as centre-of-balance when weighing student marks within multiple administrative decisions. ......................... 238
Figure 7.1. Pragmatic model for testing effects of perceptual judgement on interpretation of average student achievement. ................................. 248
List of Abbreviations, Symbols, Nomenclature

(CP) – refers to C. S. Peirce's *Collected Papers*

(EP) – refers to C. S. Peirce’s *Essential Writings*

**Inputs** – those materials, ideas or resources which enter an organization in the form of statistically-represented demands and supports (Easton, 1957)

**In-school administrators** – a category that includes both school principals and vice-principals

**Management** – those tasks, responsibilities, and practices which further organizational goals (Drucker, 2009)

(MS) – refers to C. S. Peirce’s *Unpublished Manuscripts*

(NEM) – refers to C. S. Peirce’s *New Elements of Mathematics*

**Organization** – any group of individuals who are statistically represented as such through inputs, withinputs, and outputs (Easton, 1957)

**Outputs** – authoritative decisions produced by organizations as consequents which are statistically represented as such (Easton, 1957)

**p-value** – an inferential statistic which assigns numbers to degrees of probability or chance to an event

(RLT) – refers to C. S. Peirce’s Cambridge lectures on *Reasoning and the Logic of Things*

**School leader** – an in-school administrator who actively manages the meaning of statistics before various audiences (Smircich & Morgan, 1982)

**Sign** (*upper case S*) – a representation of reality

**sign** (*lower case s*) – an irreducible triadic relation between a Sign as representation, the object that is represented, and an interpretant by any reader

(W) – refers to C. S. Peirce’s *Chronological Writings*

**Withinputs** – those processes of demand and support within an organization where members interact with statistically-represented inputs to produce a statistically-represented output (Easton, 1957)
The inquirer begins with pondering these phenomena in all their aspects, in the search of some point of view whence all the wonder shall be resolved. At length a conjecture arises that furnishes a possible Explanation, as premises. On account of this explanation, the inquirer is led to regard this conjecture, or hypothesis, with favour. As I phrase it, he provisionally holds it to be plausible; this acceptance ranges, in different cases—and reasonably so, from a mere expression of it in the interrogative mood, as a question meriting attention and reply, up through all appraisals of Plausibility, to uncontrollable inclination to believe.

Charles Sanders Peirce

Chapter 1: Introduction and Inception of the Study

People sometimes ask why I am interested in understanding the various ways that people interpret numeric information. My answer is fairly straightforward. My belief, and therefore bias, is that the professionalization of teaching and school administration in Canada is abetted and not eroded by “scientific” inquiry, provided we construe science in very broad terms and see statistics as the quintessential product of empirical thinking. Teachers and principals as a professional group can interact with and use statistics—whether the numbers are derived from assessment or from research—to advance their status, rather than becoming diminished to mere technicians in the social order. The central requirement for the professionalization of teaching to proceed is for educators to accommodate, author, re-author and interpret quantitative information in the public and professional spheres. I would not go so far as endorsing Popham’s (2004) claim that educators’ failure to address assessment issues portends professional suicide. However, I do believe an outcomes orientation in policy, decision making, practice, curriculum, and law (Jacobs, 2008) requires a profound shift in many key educational concepts.

1.1 Background of the Inquirer

I embark on this research as a former secondary school teacher of literature, languages, and history, as well as a former school vice-principal. My undergraduate studies and early graduate courses were in Canadian and European intellectual history
and French as a second language. My previous graduate work focused on the constituents of another key measurement concept—a criterion standard and the judgements needed in its construction. Over the last 15 years, I have served as a middle manager of assessment and evaluation programs in the civil services of three provinces. During that time, Ministries of Education have produced and distributed hundreds of thousands of reports that depict provincial, school division, and school performance in terms of averages. My interest is in gaining and fostering a better understanding of the way school principals interpret the averages in such reports and use these statistics in problem-solving, because principals are key figures in advancing educational practices within schools.

My own experience in school administration was acquired before large volumes of statistics about school outcomes were either available or deemed pedagogically relevant. As a vice-principal of a large grades 8-12 school, I did scan quickly the classroom marks submitted to the office at mid and end of term, to note those students who fell below a 50 percent criterion threshold and to forewarn their parents of potential consequences at semester end. Marks over 80 percent were also noted and discussed with colleagues for possible awards at the end of the year. I suspect that in the 1990s, few administrators calculated and contemplated average student scores by gender, ethnicity or geography to look for inequities. It remains an open question whether they do so now, even if every province now distributes such scores, either from classroom or large-scale assessment, in relation to some normative standard.

A couple of admissions are in order. First of all, I do not have deep expertise in mathematics or the natural sciences, although I have taken advanced statistics classes and published quantitative research. At first, having studied only high school-level algebra seemed a serious limitation. However, upon reflection, it now strikes me as an asset for
research into the practical significance of statistics, having avoided some degree of indoctrination and enabling me to see assumptions and qualities that others may take for granted. Second, as crucial as I see the “average” as being in educational matters, I have never done hands-on calculations that would define the official average student performance for an entire school, school division, or province. Nevertheless, I have generated thousands of averages as a classroom teacher and university lecturer. As a government middle manager, I have overseen the process and have many times authorized the algorithms that generated descriptive statistics for a province and for a country. Although I have filled several administrative roles in the education system at multiple levels, and have served as a vice-principal, I have never actually sat in the school principal’s chair and assumed the legal responsibility (The Education Act, 1995) for determining (in consultation with classroom teachers) a student’s mark.

Nevertheless, I have had many experiences helping senior policy makers, politicians and educators interpret data, and have witnessed their difficulties when making meaning from numbers. Working within government gave me “hands-on” familiarity with issues in data reading and use of assessment results. However, policy making—the translation of politics into principles for program development—is an arena quite different from educational practice. Government is marked by scaled-down, immediate and attainable goals, concern for unforeseen consequences, and transitory plans (Levin, 2005). A minister’s office is not the realm of utopian aspirations, long-term designs, and conspiratorial plots (Anderson, MacDonald & Sinnemann, 2004). The anecdotal usually prevails over the empirical, and budgetary considerations prevail over principles and programs. In policy circles, decisions involving measurement and assessment products often “accrete” (Weiss, 1980a) rather than “are made.” That is, without conscious reflection, different people in different offices do their jobs, and over
time a decision emerges which is an accumulation of several small choices or indecisions that build or break off from the previous one. Most policymakers—politicians, senior civil servants, the executives of stakeholder groups, and parental coalitions—sincerely want the best for youth. So the civil servant’s challenge is in reconciling disparate wants. In other terms, government, unlike the university, is practical rather than theoretical in its orientation; in many ways, government is much less idealistic than are classroom teachers or school administrators.

When I think about statistics and action, I often recall an informal conversation with an executive member of the Saskatchewan School Boards Association. He was frustrated by the difficulties in persuading educational leaders to consider statistics in their practice. “You know Darryl, when I read a geography textbook,” he said, “it tells me that about 70 percent of the earth’s surface is covered with water. Yet when I look down at my feet planted on the ground nearly every day of the year, I have a hard time grasping that idea.” A statistic, he was saying, must have meaning in our everyday lives before it is conceivable, deemed acceptable, and hence becomes actionable.

Many might consider my own view of an average as simple, even simplistic. My image is often in non-Cartesian, non-formulaic, three-dimensional, and topographical terms—as the peak of varying and shifting piles of sand—each granule representing some smaller entity, whether of an individual, a budget, or a group performance. Averages are susceptible to the laws of social gravity, constantly shifting in the policy, program and personality winds, and prone to distortion under the surveyor’s tread. Each average has a different shape. Each is capable of comparison in the moment, to reveal differences in the educational geography. However, contrary to Statistics Canada (2010), I do not hold that the meaning of any average arises automatically from the reference points intrinsic within the quantitative or graphic display (p. 5), but rather from those experiences
brought to it by the leader as reader within an organizational context. For me, averages are inevitably coloured by subjective values, while simultaneously expressive of relatively objective measurements.

Of course, expectations are difficult to disentangle from statistics. I have been challenged many times by the normative assumptions that both measurement specialists and scholars bring to questions of statistics, state power, and testing. I have participated in many debates during my career with psychometricians about the assumptions they bring to their work on significance testing. I have always preferred to use the principles of criterion-referenced measurement within assessment, because I believe the consequences of applying these principles for curriculum, teaching, and public engagement are more desirable or constructive in a policy and pedagogical sense than the normative assumptions brought to so much statistical endeavour. It is fair to say that in my work for the governments of three provinces over 15 years, I was part of a team that introduced many of the principles of criterion-referenced measurement to the national School Achievement Indicators Program, the Pan Canadian Assessment Program, and several provincial testing programs.

At the same time, I was often dismayed by the way so many policy makers, educators, and scholars were transfixed by the new large-scale testing superstructures that emerged on the educational horizon over the past decade or two, while they ignored the bog land that is classroom assessment. The highest-stakes testing in terms of students’ lives and destinies in every Canadian province lies in those tests constructed and administered by teachers. The resulting grades are considered in nearly all promotion, graduation and scholarship decisions, and in deciding whether a student proceeds to postsecondary education or enters the labour force directly after high school. Classroom grades or report card marks are also, arguably, the most labour-intensive forms of inquiry
found in Canadian schools. Classroom grades and report card marks are also a crucial sorting mechanism in social stratification: report card grades are the sole bases for promotion and retention decisions in every Canadian school until Grade 10; at the high school level, ministry examinations count for only a portion of the marks necessary for credits in key subjects toward graduation. Statistical significance may be a preoccupation for the measurement specialist in large-scale assessment, but, from my perspective, the practical significance of classroom marks is a major policy issue in Canadian education.

1.2 Beliefs of the Inquirer

Three other beliefs underwrite this dissertation. The first belief is that clarity (but not consistency) in our conceptions of an average is integral to cultivating what is called “assessment literacy.” Assessment is often framed as a relatively technical process for collecting data, divorced from the messier and more difficult process of interpreting and acting upon that information. The application of theories and experience to data is what makes assessment a technology. It is hard to go to a teachers’ conference today without seeing a discombobulated workshop on assessment as technique or method. However, few workshops address basic arithmetic or the statistical concepts that should be part of every teacher’s education.

A second belief is that greater precision about what we mean by “normal,” in and about a school and within education generally, may help both policy-makers and educators. Understandably and rightly, educators have been increasingly sensitive to the excluded and the culturally diverse in our schools. Many educational debates implicitly revolve around alternate conceptions of the normal and the exceptional. In setting out on this inquiry, I reject the claim that discussing what we mean by the average and how we represent it, inexorably leads to a discourse which standardizes, straightjackets, and excludes. I assume that our notions of the outlier, the abnormal, or the atypical should be
based on a more exacting conception of the normal. Targeting resources for the extraordinary becomes easier if we first clarify notions of the typical. In short, my depiction of an average does not imply its endorsement as a goal.

A third belief—perhaps incongruously for someone whose teaching and public administrative career revolved around language pedagogy—stems from my repeated observation that many Canadian educators have severe problems with numeracy. Notwithstanding a frequent claim they do not “do math” at work, my observation has been that most educators actually do so in multiple situations. Of course, by math, they often mean answering problems in a textbook, rather than applying spatial and probabilistic notions to daily situations. Teacher preparation programs should recognize that. We must also combat the sentiment that it is acceptable to be bad at math. Turning out repeated cohorts of teachers imbued with the zeal to inculcate early learning literacy—without a parallel attention to numeracy—impoverishes both. Literacy and numeracy interact in a ladder of conceptual development (Berch, Foley, Hill & Ryan, 1999; Schleppegrell, 2007).

So, my intent at the outset is not to dismiss but rather to clarify those elementary devices through which we seek some degree of certainty in our lives. Nor is my goal to undermine the foundations of biostatistics that have been created over the past century: discourse rules have been established and communicated for appropriate measurement and analyses in diverse situations. My goal is to examine the way that school principals interpret the average and use it to inform their everyday decisions. While this is only one aspect of decision making, the average is nevertheless an important foundation for action.

Many arguments have been made against the intrusion of statistics in educational administration in Canada. One line of organizational argument (Stirling, 2008) revolves around the maxim that technical expertise should be on tap, not on top. Administrators’
minds should not become clouded with reams of numbers, so they can think more broadly across a large horizon to contemplate alternative directions. Far better to turn to the technician for the number crunching, so the argument goes. The problem with this line of thinking, it seems to me, is its presumption that numbers cloud rather than condense, clarify, and synthesize, enabling a broader picture to be viewed more distinctly. I have had many experiences with senior educational administrators who are unwilling or unable to read statistics, leaving this task to others. I believe that this act of delegation, sometimes a deferral but certainly deference to the technocrat, ultimately diminishes the role of the decision maker. It renders the administrator dependant on the technical choices of those who are more numerate, and diminishes the authority of the administrator who has legislated responsibility for decision making.

Another argument against the intrusion of statistics in educational administration is that test statistics such as averages are somehow representative of a hyper-rationality (Wise, 1977) which rigidifies administrative thinking (Davis, 2004). One aspect of this argument is that it is not the statistics themselves, but rather the testing regimes and political agendas they serve, that are hyper-rational. Statistics are considered to promote sterility and organizational paralysis, leaving administrators unable to think laterally or innovatively. The problem with the hyper-rationalization hypothesis is its presumption that computation and calculation preordain the subsequent inference. As I will demonstrate in this dissertation, logicians have pointed out alternate ways of thinking besides the inductive and deductive, particularly the abductive (Prawat, 1999).

A more dubious set of claims against the intrusion of numbers in educational administration seems more ideological: numbers somehow represent the intrusion of industrial and market ideas in education, symbolizing colonization mechanisms at work. To stay ensconced in a qualitative paradigm is to uphold a warmer and more humane
outlook. That is, words are pliable in upholding humanistic and civic virtues, whereas budget numbers and output measures standardize and leave a chill in the educator’s heart. Moreover, numbers are considered to fragment and fracture. As Hofstadter (1985) once noted in an anthology of his *Scientific American* contributions, “Only the fool would take trouble to verify that this sentence was composed of ten a’s, three b’s, four c’s, four d’s, forty-six e’s, sixteen f’s, four g’s, thirteen h’s, fifteen i’s, two k’s, nine l’s, four m’s, twenty-five n’s, twenty-four o’s, five p’s, sixteen r’s, forty-one s’s, thirty-seven t’s, ten u’s, eight v’s, eight w’s, four x’s, eleven y’s, twenty-seven commas, twenty-three apostrophes, seven hyphens and, last but not least, a single !” (p. 27). Yet Hofstadter’s playful pangram invites its own undoing. Both the languages of words and numbers can be used to depict micro-and macro-worlds, if we understand their respective semantics. Much depends on the purposes of the author.

Indeed, when looking at the process of reading statistics at its most basic, we have only two figures meeting face to face: a person on one side and a number on the other. Yet multiple layers of metaphor, myth, and imputation of motive have masked an obvious dichotomy between the two. One is an animate creature with a beating cardiovascular system, a set of values, and a mind of its own. The other is inanimate with no palpitating heart, no intrinsic value, and no ability to leap off the page in a calculated act of free will. Nevertheless, some educational researchers demonize a number as personifying a nefarious plot. Statistics lie, they say, and liars use statistics. Statistics, they claim, tell an impoverished or thin story compared to rich and thick prose.

So too does the party who calculates succumb frequently to personification, perhaps as a way of distancing oneself from his or her work. The mathematically-aware sometimes anthropomorphize a statistic as being a person with the tears washed away. Statistics “demonstrate” a trend; the data “demand” an explanation; the numbers “tell us”
a story. All such characterizations disguise a fundamental distinction: school principals have agency, whereas a numeric sign does not. The corporeal distinction between the human figure who calculates and the numeric figure issuing from a calculation is often forgotten. My interest is in empowering the reader to reconfigure the work of the statistical author—to recognize that any reader of statistics, possessing even an elementary level of arithmetic competence, can bend a number onto another trajectory. Within any democratic community, the individual retains some degree of freedom to restructure his or her statistical environment. The general question that has nagged me revolves around how school administrators express their agency with numbers.

Thus, the central issue in this study echoes C. P. Snow’s (1959) assertion, articulated in his lecture *The Two Cultures*, that western society is split into two distinct and irreconcilable cultures—the sciences and the humanities. Those with a qualitative proclivity often presume numbers are manacles on thinking, fostering a mono-logical and lockstep linear mind. They believe that reading and acting upon numbers inexorably leads individuals and groups to a predetermined destination, whereas qualitative inquiry through textual and pictorial representation fosters more divergent and supple thinking and opens the individual to expression and emotion. To uphold a qualitative paradigm is to uphold civic virtues and creativity. In contrast, those who underline the importance of numbers in education and administration believe that only statistics enable rational and clear-headed thinking. Mathematicians and natural scientists believe their work has led to the discovery of a sophisticated and complex world, even though they sometimes metonymically assume that a statistic *is* rather than *represents* a state of affairs (Hunter, 1996).

Statistics as the application of mathematical ideas to social scientific inquiry is, at bottom, the study of variation and distribution. For the statistician, a numeric
representation requires a rigorous logic to document and describe plurality. Because numbers reflect patterns of diversity, systematic quantitative study enables debate and perhaps shared understanding. But for me, a key question has remained personally vexing: does the inferencing inherent within the construction of a number invite convergent or divergent thinking on the part of the reader?

1.3 Beginnings of the Inquiry

Of course, when developing policy relating to assessment and evaluation, I was repeatedly confronted with the intersecting challenges of measurement and values. Two decades of hands-on experience conceiving and implementing indicators systems, both provincially and federally, caused me to question how a school principal can conceptually reconcile the demands of measurement consistency and uniformity with a respect for diversity in a democratically-governed education system. So my initial path of investigation went from indicators as signs, to a theory of signs, and ultimately to the writings of Charles Sanders Peirce (Brent, 1998). Peirce appeals because he addresses many issues of interpretation compellingly, because his pragmatism accords with mine, and because of my interest in the history of ideas rather than the histrionics of ideology. William James as psychologist and John Dewey as educator were also pragmatists, and friends of Peirce, but not nearly as knowledgeable about mathematics and statistics. Peirce’s ideas on statistics, meaning, and practice seemed more germane to this study than the ideas of either Dewey or James. Moreover, Peirce always stressed the interrelationship between meaning and methods of inquiry, leading me to adopt his conceptualizations for both.

Alternative theories of interpretation are on offer. Many theories couple inference with particular statistical techniques (Huck, 2008), striving to bind the reader into the
deductive reasoning of mathematics and the inductive reasoning of formal science. Other theories deal with specific types of representations that have mathematized their objects: Wood and Fels’ (1992) model revolves around map reading but, in my opinion, is too geo-spatial in its concerns. Roth’s (2003) theory revolves around graphs, but inappropriately conflates European with Peircean semiotic notions in many crucial respects. In contrast, Peirce’s theory aligns inference and cognition with a wide variety of representations in general practice. Flexibility in application, coherence in approach to causation, a holistic view of the reader, consistency in methodological and philosophic propositions, and relevance for both statistical and linguistic representation were my criteria of choice. Peirce’s formulation makes sense to me, and helps me understand notions of “significance” in everyday activity.

Peirce wrestled with terminology, often inventing new terms, each of which he believed had ethical implications. He did not sharply distinguish between the terms “concepts,” “ideas,” “signs,” and “significance” because all these terms were closely related in his pragmatic theory. Accordingly, for this project, I hold these four terms as synonyms. In contrast, “conceptions,” “inferencing,” and “interpretation” were dynamic processes within his outlook. Likewise, although he carefully differentiated between “consequents” and “effects,” and between “thought” and “action,” these distinctions cannot be made without understanding Peirce’s philosophic precepts which I apply in the latter half of this dissertation.

So for some time, I’ve cast about for an appropriate entry point that would enable a manageable inquiry into the interplay between practical and statistical significance, issues of normative power and educator agency, and school principals’ practice with numbers. For a while, my line of inquiry followed well-worn scholarly ruts in the path toward utility (Clavijo, Fleming, Hoerman, Toal & Johnson, 2005). Frameworks of data
usefulness date back at least to the 1960s in the program evaluation literature; Michael Quinn Patton’s (2008) book has been one milestone on that route. It now seems to me that this approach overlooks as much as it illuminates my problem: it over-privileges the author of the statistic and deprecates the role of the reader or the respondent audience.

The other challenge facing me was to dramatically narrow the vast field of statistics and data to a particularly revealing concept. Focusing on the arithmetic mean seemed too constrictive, but its more promising parent, the “average,” seemed accessible from both a qualitative and quantitative perspective. An inquiry into what school principals conceive as “the average” promises provisional answers to some of my questions about diversity within apparent unity. The average as expressed verbally, but also mathematically and statistically, could yield insight into how readers of numbers go about the act of interpretation. The average seems central to the appropriate reading and interpretation of most quantitative research by adults: grade point averages are one of the most frequent measures in psychological research. The average is also often used by the media when describing trends. Computations of the average, moreover, extend across educational finance, educational research, and educational assessment.

Many researchers have argued that standardized testing tends to render organizational structures more brittle, that the mass media misuses data, that think tanks distort numbers, that large-scale testing is an appendage of state power, and that creating accounts for accountability purposes are an atavism in education (Graham & Neu, 2004). I wondered whether this line of thought stems from the quantity of statistics generated, or from the innumeracy of those who are contemplating statistics. Therefore, my purposes here stem from doubts arising from claims that statistics, somehow, suborn school principals, as readers of those statistics, into a uniform interpretation of the world, erasing principals’ agency in decision-making. My initial belief is that even deceptively simple
I have another, overarching research agenda too. Educational administration is missing much in the realm of mathematical cognition, both in the role of statistics in educational administration, and in the way administrators think mathematically in their practice. Management through mathematics may be studied as a distinct field, separate from leadership through the motivation of manpower, the manipulation of metaphor, the allocation of materiel, and the consideration of moral issues. For a variety of reasons, educational administration is often considered to be primarily managerial or regulatory in orientation, predisposed to maintain an organizational or community balance. In contrast, leadership (in my opinion) involves the deliberate attempt to foster a controlled imbalance. My presumption is that statistics can support both managerial and leadership functions, precisely because they are susceptible to so many divergent and even fractious interpretations in the public domain, potentially accentuating conflict rather than reducing profound socioeconomic inequities. In short, my conjecture is that school principals’ conceptions of “the average” can reveal much about their normative judgement, their sense of agency as leaders, and statistical inference in practice. A study of the concept of the average also enables me to continue a research interest in the practical significance of particular measurement concepts. Measurement and mathematical ideas do not arrive in tablet form from a Platonic heaven (Lakoff & Núñez, 2000), but rather consist of ideas held by humans with their feet on the ground. Thus, for this study, I view the “average” in practical terms: our understanding of its meaning and derivation is values-laden,
sometimes bears an emotional charge, refracts and reflects diverse ideas, and may be contested in leadership.

Accordingly, drawing on beliefs for plausible explanation, my presuppositions when undertaking this study are:

1. Our conceptions of an average as a statistic are central to comprehension and use of much quantitative research, classroom and large-scale assessment results, budgetary outlays, and other educational phenomena.

2. Our conceptions of an average as descriptions are important for reciprocally bringing definition to those qualities we consider to be exceptional, excellent, extraordinary, diverse, atypical, or abnormal.

3. An average is an umbrella term for a host of representations which are authored by humans, and which can be expressed numerically, graphically, or verbally.

4. Averages are often used to calculate, compare and contemplate student achievements in Canadian schools.

5. The significance of any average as a symbol for school administrators is conditional or contingent on the experiences that any reader brings to the number, including my own experiences as researcher.

6. An individual’s understanding of an average may reveal much about that person’s normative judgement and sense of agency, while also shaping the organization within which they work.

7. Conception, as a cognitive act, refers to the substantive features of interpretation. Inference, as a reasoning process with concepts, is central to interpretation.

8. Interpreting basic statistics is an important aspect of the school principal’s job description and supports the principal’s managerial and leadership roles.
Chapter 2: The Interpretation of Average Student Achievement

In this chapter, one facet of educational leadership is described as crucial for educational improvement—the management of statistical meaning. The purposes of the study are outlined and the importance of the statistical concept of the average is considered within different orientations to instructional leadership and alternate models for decision-making with data.

2.1 Management and Meaning

What is called “the management of meaning” has been recognized for several decades as an element of successful leadership across multiple settings (Angus, 1996; Magala, 2009). A leader must be a social architect who understands the organization, shapes symbolic life within and outside it, and thus changes mental maps. Leadership itself, according to Smircich and Morgan (1982), “is realized in a process whereby one or more individuals succeed in framing and defining the reality of others.” (p. 258). This process has been characterized as a socially-interactive relationship wherein those who lead acquire or have the power to verbally or pictorially shape the reality of others. Leadership means that followers surrender their authority to interpret events and thereby to make organizational activity sensible. Within the process of managing meaning, issues of power arise and conflict often ensues. Therefore, formal roles are questioned, and

---

1 For the purposes of this inquiry, a leader is defined as an in-school administrator who actively manages the meaning of statistics before various audiences (Smircich & Morgan, 1982). Administrators include both principals and vice-principals. Management includes those tasks, responsibilities, and practices which further organizational goals (Drucker, 2009). An organization is any group of individuals who are statistically represented as such through inputs, withinputs, and outputs (Easton, 1957).
informal roles are thereby clarified within the organization. How leaders help others make sense of events inside and outside an organization enables leaders to give direction, mobilizing followers and immobilizing critics, gluing members together in common cause, and rendering work meaningful (Lips-Wiersma & Morris, 2009).

The central premise in many theories of organization and management is that underlying images or verbal analogies lead us to conceive and plan for organizational events or situations (Morgan, 1997). As originally theorized, the leader as figure initially defines a focus of attention, through his or her actions, to bracket an element of experience as ground. Then followers interpret the leader’s action within a wider context wherein an event’s significance both resides and is ascertained, either accepting the leader’s gloss or asserting their own interpretations. Thereafter, organizational actions become holistically meaningful in a negotiated process such that followers link the leadership figure’s assertions or actions with the experiential ground (Smircich & Morgan, 1982).

However, it seems to me that this approach to the management of meaning suffers from several shortcomings. First, it assumes that leadership action precedes interpretation—that is, leaders “do” before they make any sense of a phenomenon themselves. Second, while scholars have extensively described managers’ use of metaphor and verbal analogy for defining reality in alternative images of organization, they have not looked at statistics as vehicles for accomplishing (dis)similar leadership ends in education. Only recently have educational researchers (Anderson, Leithwood, & Strauss, 2010) recognized that school principals’ interpretation of data has both a pre-cursive and recursive role in guiding use. Third, little research has been conducted into how leaders—both autonomously and in concert with others—actually reason with and interpret various kinds of statistics in practice. Fourth, researchers often assume that
managers lead through “assertion” when there may be alternate ways of managing meaning. Fifth, our notions for the terms “meaning” and “significance” remain remarkably ambiguous, notwithstanding a century of psychometric effort at definition. Sixth, in my estimation, scholars of the management of meaning have over-equated statistics with machine images and formal scientific metaphors while also caricaturing numbers as incompatible with living organisms, cerebral information-processors, political systems, cultural formations, or other images of organization (Morgan, 1997). And most important, we know remarkably little about the actual concepts that leaders have for seemingly obvious arithmetic ideas. Presumably, leaders have their own statistical ideas before or as they manage the meaning of statistics for others, but this remains largely terra incognita in educational research or, indeed, in many realms of management study.

Mapping this educational terrain becomes pertinent because the general Canadian public is becoming increasingly aware of statistics about student achievement. This information is often presented in terms of averages. School principals, as key figures in Canadian communities, are expected to interpret and act upon such outcomes information to improve teaching and learning. Principals’ conceptions of "average student achievement" are therefore important when planning and evaluating, and when fulfilling their legislated duty to “establish, in consultation with the staff, the procedures and standards to be applied in evaluation of the progress of pupils and in making promotions” (The Education Act, 1995, 175(h) (k)) in Saskatchewan, similar to other provinces.

Current models for investigating the impact of statistics in educational administration customarily revolve around the statistic’s utility for administrators in the probabilistic pursuit of maximum gain. Indeed, recent inquiries (Council of Ministers of Education, Canada, 2012) into Canadian school principals’ views about the flood of outcomes data entering Canadian schools have continued to frame questions in terms of
“data use.” Surveys revealed that principals held negative perceptions about the ease with which national and international assessment results may be obtained, interpreted and used in making instructional changes or constructing action plans (p. 163). Only eight percent of these in-school administrators reported using this information frequently or sometimes.

Nevertheless, Canadian principals were positive about using results with teacher groups and at staff meetings, about discussing these results with parents and guardians in the school community, and about their teachers using test results to make changes in instruction. Although school principals were more optimistic about the availability and ease of interpretation for provincial assessment results, only half said they used this information for improving instruction. In contrast, approximately 95 percent of Canadian school administrators reported using teachers’ classroom assessment results for multiple purposes. In general, the Council of Ministers of Education, Canada’s (CMEC’s) surveys corroborated an earlier, cross-country study which found that school principals were more optimistic and felt they had a greater margin for maneuver than did teachers within a policy context which emphasizes student outcomes (Lessard, Kamanzi, & Larochelle, 2008). Overall, a complex picture has emerged about school principals’ uses, purposes, and perceptions of data, illuminating wide differences within and among the provinces.

The vice-principal’s place in this Canadian picture remains obscure (Cooper & Levin, 2010). Although principals’ work has been characterized as “fragmented, fast-paced, and varied; it involves long hours and a relentless workload, along with demands from multiple, diverse stakeholders” (Spillane & Lee, 2013, p.1), the vice-principal’s role in data interpretation or use has not yet been extensively investigated in Canada (Cousins, Goh & Clark, 2006; Goh, Cousins & Elliott, 2006). Some American researchers have started to delineate the vice-principals’ functions from those of the principal in data-based
decision-making. Vice principals’ specific tasks are often delegated by the principal (Wong, 2009). Because vice principals do not assume the responsibilities and hence stresses of the full “burden of command” (Allison, 1997; Spillane & Lee, 2013), because they often are tasked to monitor attendance through statistics, because they typically have less experience in leadership positions, and because—in my experience—they often operate as “consigliere” in supplying information to principals who “front the organization” before various audiences (Armstrong, 2010)—it might be speculated they have more time and indeed the responsibility to generate and contemplate statistics within a school. How we model the division of managerial responsibility will in turn determine any model of data use.

However, data use models may disguise as much as they reveal about the management of meaning. Umbrella terms such as “use” overlook dozens of more precise terms in a thesaurus. Frameworks for statistical use sometimes imply that school principals as readers of statistics must be instruments of others’ (authors of statistics’) rationality, denying the reader’s agency and upholding the author’s reasoning as correct. For example, an inferential statistic may be significant and generalizable from the statistician’s perspective, but not from the “user’s” point of view; conversely, the “user” may inferentially find a descriptive statistic to be more significant and have more general import. Moreover, we have at least three decades of research in behavioural economics which convincingly demonstrates that adults often do not think and operate on the basis of economists’ strictures about utility. Instead, adults rely on experience and intuition, however defined. Most crucially—unless we adopt Pavlovian precepts—most would argue statistical interpretation comes before or as we act, not afterwards.

Statistics are one type of sign or representation. Although there are several generic theories for the interpretation of signs, most revolve around the interpretation of
linguistic or graphic material. Perhaps surprisingly, there are few theories about practitioners’ interpretation of statistics as signs, except for theories tied to the specific techniques and inferences of a statistical author. The most comprehensive and cohesive theory we have for sign-interpretation by a statistical reader, in contrast, derives from the philosophy of Charles Sanders Peirce. No other set of propositions has such a firm footing in mathematics, nor has such a solid pedigree in North American education (through Dewey) and psychology (through William James). Moreover, his theory is integrated with a theory of reasoning which embraces the specialist and non-specialist alike.

2.2 Research Purpose and Questions

This dissertation tests the applicability of Peirce’s theory of interpretation, using a mixture of methods that he advocated (phenomenological, randomized experimental, and historical) to study school principals’ conceptions for a pivotal statistical concept—the average. In looking at the “fit” of Peirce’s theory, I address several issues in cognitive styles and cognition that scholars have identified when attempting to explain adults’ contradictory responses to basic arithmetic ideas. In other words, I bring a different lens to some of the “biases” that cognitive psychologists have recurrently identified in judgements under uncertainty (Hilbert, 2007), offering alternate, pragmatic explanations.

Although Peirce’s pragmatic-semiotic theory of interpretation is situated within a philosophy of practice, its application is far from straightforward. His theory has not heretofore been applied in a study of administrators’ interpretations for numeric signs. Nor has Peirce’s phenomenology as a research method been previously adopted to investigate the ways people interpret the significance of performance statistics. Although Peirce’s ideas are substantively different from European theories of sign-reading which
have gained acceptance in a few academic circles, some scholars have conflated his distinctly North American theory with others’ theories. Moreover, Peirce’s original ideas have been obscured by disarray in his papers, through obscure terminology, by philosophic abstraction, as well as by scholars’ inability to grasp some of his mathematical and ontological premises.

Notwithstanding these challenges, Peirce’s writings enable us to discern the substantive differences between the practical significance of a statistic for leading change versus the statistical significance of a change in leadership practice, with alternate types of inferencing brought to each. Thus, the research zone in this study is not school administrators’ data use, but earlier, at the point where they receive and read a statistical report. My overall goal in examining these issues is to advance the thinking of educational administrators and scholars about instructional improvement in Saskatchewan and elsewhere. I also implicitly uphold numeracy as an important goal in public education, which school principals can and do model before their communities.

In short, my research problem is:

**How significant are statistical representations of “average student achievement” for school administrators as instructional leaders?**

To address this question, I pose two inter-related research questions. One question emphasizes the substantive ideas that school principals have for the average as a quality; the other focuses on administrators’ cognitive styles as interpretive processes brought to an average as a quantity.

1. **What are school administrators’ conceptions for the term “average” student achievement?**

   My related conjectures are:

   a. Given that Canadian school principals report being more optimistic than teachers about large-scale assessments and the normative assumptions upon which these assessments are based, and that school principals feel more room to maneuver in their practice than teachers, we could infer that Saskatchewan school principals’ conceptions of an average are nuanced and multiple.
b. Since all school principals were once teachers, and since they continue to work in educational settings with teachers, we could infer that they will compute an average as do most teachers—as an arithmetic mean weighted in relation to curriculum priorities.

c. Because all school principals have legal responsibility for school marks, and because their previous experience as teachers will predispose them to consider student achievement in relation to classroom marks, we could infer that principals will frame the consequences of averages in terms of classroom assessment rather than large-scale assessment.

d. Because all school principals are practitioners and because relatively few have training in mathematics and statistics, they could discount notions of statistical significance as the comparison of two means.

2. How does framing “average” student achievement, in the positive and as a statistic made public, influence school administrators’ conceptions of the average and their problem-solving approaches?

The associated hypotheses are:

a. School administrators will adopt more rational and analytical problem-solving dispositions when school averages are framed negatively than when they are framed positively.

b. School administrators will adopt more experiential and intuitive problem-solving approaches when averages are framed as confidential to the school, than when they are framed as in the public domain.

c. Vice-principals will have more rational and analytical pre-dispositions to the interpretation of average school performance than will principals.

d. Experience as an educator and as an administrator, position, gender, education level and statistical self-efficacy will not substantially predict school administrators’ problem-solving pre-dispositions when reading statistical averages.

2.3 Contexts for Considering the Average

The management of statistical meaning is important because closing gaps in achievement among groups of students is a goal of educational policy-makers in Saskatchewan, North America and other jurisdictions. Ministries of education and school boards are increasingly making public a variety of reports that describe student performance in terms of the “average.” This measure of centre is omnipresent in educational administration: grade point averages (Sadler & Tai, 2007), graduation rates,
average per pupil expenditures, average attendance, and average class-size are examples. Public reports, deriving from large-scale assessments and teachers’ classroom grades, often rely on averages to summarize student performance as a measure of school quality. Averages from a particular school system are often compared over time and become important reference points for instructional and school improvement. School administrators are expected to explain reports that focus on averages to staff and to their communities, and to exercise leadership in improving student performance.

In this Province, many policy-makers are concerned with student performance on (inter) national assessments. Since the 1990s, “while Saskatchewan students score similar to international standards, they score significantly below the Canadian average” (Saskatchewan Ministry of Education, May 2012, p. 1). Such comparisons within the Programme of International Student Assessment (PISA) and the Pan Canadian Assessment Program (PCAP) results, policy-makers argue, “may place Saskatchewan students at a serious disadvantage for acceptance into postsecondary education programs of study, as well as employment opportunities” (Saskatchewan Ministry of Education, May, 2012, p. 3). Underscoring the issue, some universities are reconsidering their approaches to weighing student marks for admission purposes. For example, in November 2011 the University of Saskatchewan announced it will adjust its admission policy “to level the playing field” for Alberta students, but not for applicants from other provinces. Admissions officers now recognize, for university entrance, the best average Grade 12 scores obtained from the Alberta teacher or a diploma examination or a 50-50 blend of the two, thereby appraising graduation marks from Alberta as a more consistent measure of achievement than in other provinces (Alberta Teachers’ Association, 2011, University of Saskatchewan, 2011).
Across Canada since 1981, every province has (re)introduced large-scale assessments to supply policy-makers, educators and sometimes the public with information about student performance; in several provinces, examinations have conducted without stop since the mid-century. In Ontario, the Education Quality and Accountability Office annually measures the achievements of well over 900,000 Grade 3, 6, 9 and 10 students for the purposes of educational improvement and accountability. Results for individual students are relayed to every school principal in the province, with both criterion and normative comparators. On the national level, the CMEC’s (2012) Pan Canadian Assessment Program has reverted to national averages as the point of comparison in its reports, using the normative assumptions in item response theory to calibrate its scales and results.

In the United States, the policy concern with below average student achievement has been manifest since the 1980s in No Child Left Behind legislation, the controversial federal law which mandates testing in American schools (United States Department of Education, 2010). Since 2009, the American federal government has directed billions of dollars to states for turnaround efforts that focus “on the bottom five percent” of the nation’s lowest-performing schools, requiring: school closure; conversion of a public school to operate under an independent governance model or charter; replacement of the principal and 50 percent of school staff; or transformation of a school with accompanying reforms, including replacement of the principal. Statistics thus have profound consequences for school administrators in the United States.

The Obama administration has not yet re-authorized the No Child Left Behind Act, because of Congressional deadlock. Instead, the United States Department of Education has granted waivers to 34 states, with seven others in the offing, conditional on their submitting plans to Washington. The waivers allow for continued federal funding to
states provided the state develops assessments in basic skills for all students at select grade levels, and that teacher quality is, in part, measured by students’ test scores. The Obama administration states that:

States and districts will collect and make public data relating to student academic achievement and growth in English language arts and mathematics, student academic achievement in science . . . . At the high school level, this (sic) data will also include graduation rates, college enrollment rates, and rates of college enrollment without need for remediation. All of these data must be disaggregated by race, gender, ethnicity, disability status, English Learner status, and family income. (pp. 8-9).

Since 2000, the Organization for Economic Cooperation and Development through its Program of International Student Assessment (PISA) has conducted assessments of 15-year-olds to collect data about student achievement in basic subject areas. What started as a 16-country venture has now broadened in its last cycle to include 65 countries around the world. In its sampling design, item analyses, psychometric modeling, scaling, calibration, equating and reporting of results, the PISA assessment assumes “a multivariate normal distribution” (Organization for Economic Cooperation and Development, 2009). Although Canada has fared well in international comparisons, gaps have appeared among the provinces and between gender, language and ethnic groups.

Most studies of workplace statistics show that the general public reasons quite differently from mathematicians: “problem-solving at work is characterized by a pragmatic agenda and geared to solving particular problems” (Noss, Hoyles, & Pozzi, 2002, p. 17), whereas mathematics as a field of academic inquiry and subject for instruction is frequently not focused on the practical or applied use. Cognitive psychologists and anthropologists have repeatedly found that adults do not interpret or read statistics in the ways anticipated in formal mathematics: “The failure to adhere to
these strictures leads to descriptive models of human rationality that have profound public policy implications” (Stanovich & West, 2008, p. 690).

### 2.4 Statistics and Instructional Leadership

Of course, the management of statistical meaning is predicated on particular views about the manager’s role. The archetypical school principal as statistical reader reflects an ideal for that occupational position, particularly as it relates to instructional change. Instructional leadership emerged as a new and amorphous construct in American educational administrative studies in the wake of and as refutation to the Coleman Report (Coleman, 1966) that suggested schools could not substantially improve student learning because of the socioeconomic conditions that prevailed in students’ homes and communities. Forty years later, “instructional leadership” is touted in several forms, and is widely-accepted by policy-makers and practitioners as an essential element of leadership practice in schools (Hallinger, 2005). Indeed, recent reviews largely confirm early assertions about leadership and student learning (see Leithwood, Harris, & Strauss, 2010; Robinson, Hohepa, & Lloyd, 2009; Waters, Marzano, & McNulty, 2003)—the former does have a “significant” impact on the latter. Findings in the last ten years are fairly consistent: leadership can elevate student outcomes by as much as five to seven percent, but largely indirectly, and as mediated or channeled or coordinated through the efforts of others.

“Leadership for learning” is the newest label in a vast body of scholarly and advocacy literature relating to instructional leadership (Hunter, 2011). Within that literature, we can discern at least three orientations to the management of meaning. Each derives from alternate research traditions, contrasting assumptions, and differing stances toward numbers such as the average. Succinctly, if over-sharply, those holding a First-Among-Equals viewpoint focus on working with teachers to stimulate change, those
promoting a Transformational Leadership orientation concentrate on the characteristics and behaviours of leaders within schools, and those championing a Large-scale Reform orientation see leaders as implementing system-wide change. These perspectives derive from different orientations to education itself: the First-Among-Equals perspective is grounded in the literature on effective teaching and professionalization of teachers, the Transformational Leadership perspective in the literature on effective leadership, and the Large-scale Reform perspective in policy and public administrative prescriptions for systemic reform. Each casts the principal in somewhat different molds, with different stances toward statistics.

The First-Among-Equals perspective views the principalship from within a curriculum and teaching orientation, originating largely in the effective schools research of the 1980s. Insofar as the principal considers statistics, she or he will focus on teachers’ classroom assessment results. Researchers in this tradition assume that schools are dedicated to teaching and learning as classroom processes, and consider the principal as head teacher or First-Among-Equals. The central and often unstated presupposition in this view of leadership is that the principal’s central task is supporting teachers in their role. The teacher-student relationship is the focus and it is assumed that enhancing the capacity of the teacher will improve the student’s learning (Matthews & Crow, 2003). Instructional leadership is largely about promoting teachers’ professional growth; the principal is not an instructional leader, but rather the coordinator of teachers as instructional leaders. For Blase and Blase (2004), instructional leadership is the principal working cooperatively with teachers in a form of “mutual nudging in the profoundly cooperative search for answers” (p. 4) to instructional problems.

For this reason, advocates of the First-Among-Equals paradigm recoil against the intrusions of external bureaucracies and especially against externally-driven assessment.
“Power-oriented administrators wielding reductionist algorithms for guiding teaching,” Blase and Blase (2004) contend, “have all too often prevailed in learning matters.” (p. xiii). They have barely concealed scorn for those like Leithwood (2006) who has distilled teaching into a deceptively simple formula: \( P_j = f(M_j, A_j, S_j) \) where \( P \) stands for a teacher’s performance, \( M \) stands for the teacher’s motivation, \( A \) stands for the teacher’s abilities, professional knowledge, and skills, and \( S \) represents their work settings, the features of their school, and classroom (p. 187). From the First-Among-Equals perspective, teaching is viewed as an exquisitely complicated craft (Blase & Blase, 2004, p. xx): neither the instructional leader nor a statistician should easily or unilaterally redesign that craft work as communally-shared and customary.

In contrast, those adopting a Transformational Leadership perspective often draw their beliefs from leadership studies, alternative views of organization, and strategic policy making. Proponents tend to paint the principal as a (neo)heroic leader. Although this scholarly orientation is widespread in all leadership studies, it is particularly common in the American setting (Elmore, 2000). Proponents (Sheppard, 1996) often focus on the autonomous behaviour of the leader and downplay societal influences, although of late, they have accepted distributed leadership models (Spillane, Halverson, & Diamond, 2001) and have recognized community context (Leithwood, Patten, & Jantzi, 2010) as important considerations in contingent theories of leadership. Proponents draw on theories of power and effective decision-making as central assumptions. Researchers in this vein often implicitly or explicitly address issues arising in quantitative research about the actions that effective leaders take to improve student achievement within schools and sometimes without (Mangin, 2007). Transformational researchers aim to identify those practices that leaders use to influence or shape school or community expectations as prerequisite to changing teachers’ outlooks and beliefs. Researchers in this tradition
focus on the relationship between principals (leaders) and teachers (followers), not the relationship between teachers and students. They emphasize that particular organizational prerequisites must be in place, that evidence will be drawn from a wide variety of sources, and that data define problems rather than serve as the basis for decisions (Anderson, Leithwood, & Strauss, 2010).

A third and more recent outlook on the principalship—the Large-scale Reform perspective—derives from studies of reform across schools and school divisions, and the power of statistics within large systems to inform judgement and to direct people’s attention to the core purpose of schools—student achievement. Large-scale Reformers are preoccupied with altering educational systems that tend to become hidebound. In this perspective the educational leader is often cast as strategically framing problems with data, and as encouraging teachers to use data to inform professional judgement over the long haul within the instructional program (Bellamy, Fulmer, & Muth, 2007). Rather than seeing assessment as deleterious as do those in the First-Among-Equals school, or as a technical issue that must be managed, or as a post hoc justification for leadership intentions, as do those holding Transformational Leadership perspectives, those adopting the Large-scale Reform perspective see assessment as integral to reform efforts. They foresee leaders helping various audiences interpret and act on a combination of internally-generated classroom data and externally-generated information, through feedback loops, to shape, induce and guide the education system to focus on improvement (Hume, 2010). It is not professional development alone, nor exceptionally efficacious leaders at the local level who are drivers of change. Rather, evaluative data spur and speak to moral purpose; data enables both professionals’ and administrators’ judgements to become progressively more refined and nuanced about student learning.
In the Large-scale Reform perspective, the effective leader seeks to improve working conditions for teachers as they focus on student learning and achievement, while both the principal and teacher attend to feedback cycles that provide information about student learning.

These three different perspectives on instructional leadership are further clarified by comparing their respective proponents’ views on assessment as processes, on data as assessment products, and the function of statistics as representations of what happens in schools. Those adopting the First-Among-Equals lens emphasize praise and continual positive reinforcement for the teacher as central to the professional growth that improves student learning (Blase & Blase, 2004). While those in the First-Among-Equals school do accord a role for data and feedback, they are concerned about external testing as an instrument for command and control. “Most supervisors are prepared to use only standard evaluation systems and mandated procedures that encourage them to assume a control orientation as opposed to a growth, collaborative, empowering orientation based on nonthreatening classroom observation and data gathering.” (Blase & Blase, 2004, p. 29). Proponents do recognize the value of classroom assessment and needs assessment that originate within schools, but see large-scale testing as a throwback to behaviourist models of stimulus-and-response conditioning, or as ideologically-motivated by corporate designs on the public sector.

The central problem with large-scale testing, First-Among-Equals proponents believe, is that it identifies failure that can demoralize and demotivate the learner. Large-scale testing creates normative pressures that create winners and losers. Unless used judiciously, assessment and data divide the classroom or professional collective to diminish both, and can be used by unscrupulous administrators as a form of punishment that constrains and corrodes the learning process. Large-scale assessment is an
instrument of state power and an administrative cudgel standing in the way of professional self-expression. Together with data as the statistical embodiment and over-simplification of students’ humanity, assessment negates and binds the instructional leader. Testing schemes narrow instruction, create pressures on both educators and students to “game the system,” and lead to an excessive focus on “bubble kids” (those whose average marks can be inflated, but whose hopes are easily punctured in real-life skill situations), rather than the truly disadvantaged students.

For those in the Transformational Leadership outlook, the primary source of information about effectiveness in instruction is academic research, with the averages therein. They draw on Leithwood’s (Cousins & Leithwood, 1986) interests at the beginning of his career—about the conditions and types of use to which scholarly reports are put in education. Transformationalists generally do not see large-scale assessment as furnishing forms of feedback for instructional leaders: assessment does not figure in any of Leithwood’s or Hallinger’s research designs as an independent variable, factor or mediating influence, but only as a dependent variable illustrating a leader’s effect. For example, Leithwood sees testing “with its attendant baggage” (Leithwood, Jantzi, & Mascall, 2002, p. 3) as immovable, bureaucratic, and outside the school leader’s control; research evidence and not assessment of outcomes will effect change.

Rather, Transformationalists place greater stock in school or district monitoring systems that focus on input and process indicators (Leithwood, Aitken, & Jantzi, 2006) that are periodically reviewed by professional learning communities. The one-way transfer of rigorous evidence from academic research into leadership circles and eventually to the school principal is advocated as the best way for guiding principals’ actions. Assessment is an applied set of measurement techniques that are best delegated to others as a technical issue. More often than not, Transformationalists see assessment
as creating public relations problems that distract from the core activity of schools and that create managerial headaches (Leithwood, Stienbach, & Jantzi, 2002), or as episodically supplying justifications for leadership actions. While they do see large-scale assessment data as having important sorting functions, they point to “feed forward” or “washback” effects that compel organizations to rigidify in their programming. Some point to data as potentially supplying a crutch to inept leaders without a firm values position that must be articulated in their leadership vision (Hallinger & Heck, 2002, p. 6).

For those adopting Large-scale Reform approaches, in contrast, assessment is crucial to instructional leadership because it provides feedback in cycles at the classroom level. They point to meta-analyses demonstrating that feedback falls in the top five to ten highest influences on achievement. To understand the role of assessment is therefore to appreciate the purpose, effects, and types of feedback. Large-scale Reform proponents conceive of a continuum of instruction and feedback (Hattie & Timperley, 2007). At one end of the continuum is a clear distinction between providing instruction and providing feedback to correct, which has evaluative and surveillance purposes. But when feedback and adjustment are combined in the middle of the continuum, they are so intertwined that the process itself takes on the form of new instruction, rather than informing the student solely about correctness.

At the other end of this continuum, assessment feedback is indistinguishable from the process of learning. To fulfill these instructional and learning purposes, feedback needs to provide information specifically relating to the task or process of learning. It must fill a gap between what is understood and what is to be understood. The instructional leader’s central task is to ensure that gap is filled, and the teacher’s task is to move learners along the continuum toward self-regulation.
This continuum underlies the Large-scale Reformers’ distinctions between assessment of, for and as learning. Whereas assessment of learning foresees feedback as fulfilling surveillance and accountability functions as public administration concerns, and assessment for learning as fulfilling instructional improvement purposes, assessment as learning implies that both the teacher and learner have assimilated the feedback cycles to such a degree that that they are synonymous with professional growth and student learning in themselves. The era of relying only on the best judgement of the principal for making decisions has passed. Because the public has come to expect sophisticated answers to questions such as "How well are you doing?" and ‘What has changed?" the media are interested in school affairs, and data has entered the public domain. Data will be accumulated, disclosed and used, and it is incumbent on the principal to use data wisely (Lunenburg & Irby, 2009).

In sum, how school principals should manage the meaning of a statistic will hinge on the archetype of instructional leadership that is upheld. For those emphasizing that the principal is First-Among-Equals, statistics are viewed as incarnations of power, data warehouses are repositories of a commercial interest in undermining public education, and leaders who use the data often have an intent to divide the profession. Statistics are another form of text, manufactured by conservative groups in a context that is societally determinative. On the other hand, proponents of Transformational Leadership tend to see statistics as generalizable, and context-independent within a contingent view of the environment. It is the enactment of the same basic leadership practices—not the practices as habits themselves—that are responsive to context (Leithwood, Day, Sammons, Harris, & Hopkins, 2006). While the proponents of Large-scale Reform acknowledge and recognize that assessment can be an administrative device rather than pedagogical support, most adopt constructivist approaches that consider the meaning and
hence the power as inherent not to the numbers but to the professional conversations (Earl & Katz, 2002; Earl & Timperley, 2008) and social constructions made with the statistics within varying contexts as situated or circumstantial (Creighton, 2007). At odds are fundamentally different notions of representation and of context.

Hence, we can forecast different administrative stances toward statistics such as an average. What proponents in all three perspectives do share is the fervent belief that the school leader can in some way elevate student learning and achievement above the average by appropriately managing the meaning(s) attributed to “average”? In the most recent meta-analysis of leadership effects on student achievement, Leithwood and Sun (2012) concluded that future researchers should place less emphasis on developing integrative leadership models and concentrate more on specific leadership practices that have emerged (p. 412) in light of the massive amounts of evidence at hand.

2.5 Data and Decision Making

Regardless of perspective on instructional leadership, school principals may deploy an average as evidence within a decision-making process (Levin & Datnow, 2012; Marsh, Pane, & Hamilton, 2006; Mertler, 2007) in at least four ways: to make, inform, legitimate, or promulgate a decision. Each use of the average relates to leaders’ purposes and their definitions of evidence. In the first use of an average—to make a decision—the statistic preordains choice. The evidence is deemed to precede and lead directly to the decision as an exercise in optimization. The combination of data, a costing model, and an optimization algorithm together compel a decision with minimal human intervention. School administrators’ role is reduced to making a computation. Hence, the averaging algorithm becomes crucial. Of course, the primary risk in “data-driven” decisions that rely exclusively on hard evidence deriving from algorithms and rigid decision models is that computer software may transform evidence into a decision without human
engagement at all. Statistics always provide an incomplete or misleading representation of reality.

In the second approach—as evidence to inform a decision—the average is not the basis for but rather one of several “inputs” into a choice. The other inputs include experience, bargaining with stakeholders, and general impressions (Tingling & Brydon, 2010). Expressions of power and details of specific circumstance may also enter the decision process, thereby undermining the orthodox model of rational choice. Here, the evidentiary role of an average is akin to due diligence: it either confirms or disconfirms the leaders’ initial subjective beliefs and preferences. All proceeds smoothly if the numbers confirm the leaders’ beliefs, but a dilemma arises if the statistic disconfirms the initial subjective appraisal. Decision makers must either trust the statistic, without knowing whether it is reliable or if it has been selectively shaped by those who created it. If they side with the evidence, leaders are implicitly shifting toward the “make mode;” if they side with their intuitions or with the positions of other stakeholders who may have contradictory evidence, then the statistical evidence has illuminated a choice rather than compelled a direction. Such an approach may be most useful in organizational diagnoses and strategic planning.

A third use for an average is to legitimate a decision. When used for this purpose, statistical evidence is created or modified to ratify a decision that has already been made with other inputs. This approach is distinct from decision making without formal evidence: leaders often face complex, poorly structured situations for which unambiguous evidence is simply not available, and for which they will resort to heuristic or intuitive decisions. But where evidence is available, a conflict may arise between the evidence and the decision-makers’ strongly held beliefs. If the evidence is publicly available, decision makers may simply ignore or override the evidence, at their peril, or
they may re-interpret the existing evidence so that it supports their beliefs, leading to decision biases and false consensus. The downside of ignoring disconfirming evidence is clear. The Enron scandal, the Challenger space shuttle tragedy, and any number of bankruptcy proceedings occurred, in part, because management read the evidence selectively.

The fourth use that school principals may make of an average is to promulgate a decision. If using a statistic to legitimate a decision is a weak form of decision-based evidence making, its stronger form can be discerned when leaders manufacture new evidence for implementing a decision made elsewhere without evidence. The aim is generally to signal rigour or rationality in the administrative process, thereby building confidence in their leadership. Here the evidence is deliberately created with the intent to trigger an information cascade to subordinates or to external audiences, for a herding effect (Raafat, Chater & Frith, 2009). When an average is used in this way with a well-informed audience, the internal audience will almost always perceive the evidence negatively, thereby undermining the legitimacy of the decision it was designed to support. However, it can be effective with external audiences, because the new evidential constructions reinforce leaders in their position before external constituencies, as individuals able to divine a complex and unpredictable environment. This might be characterized as using social science as a cape to reinforce the heroic role.

In all evidence-related decision processes, we need to know more about leaders’ interpretations for a statistic as evidence. However, there is little North American research about educators’ conceptions for elementary statistics, including the average—whether as mean, median, mode, standard deviation, or otherwise. We need to cut through the data-driven or data-informed mantra (Ikemoto & Marsh, 2007) to find out whether and how school principals are compelled, informed, dissuaded, or encumbered
by descriptive statistics in their practice. Much of the literature currently available about calculations of the mean dwells on methodological and procedural issues in research and assessment to ensure conclusions uphold requirements for reliability, validity and generalizability—the concerns of the statistical author. Far less attention has been paid to the processes by which statistics support routine decisions within local work contexts and to the workplace understandings that develop with this evidence. As Phillips (2007) has noted, “the complex relation between evidence and generalizations, and particularly between evidence and courses of action, tends to be oversimplified” (p. 377). Although leaders at all levels of the school system have a long tradition of collecting and using (or sometimes abusing) statistics, “evidence-based decision making” is a relatively recent phenomenon in education. The perimeters for this field of inquiry are still very much under construction (Moss, 2007).

Even a superficial review of the literature on statistics, leadership and decision making suggests that the average may be considered from many angles, as a(n):

- adjectival quality or symbolic notation of quantity;
- arithmetic representation of a norm (Gal, 1995);
- probabilistic statement of likelihood or algorithmic operation to cope with uncertainty (Stigler, 1986);
- mathematical axiom (Bibby, 1974);
- measure of centre in biostatistics and historical practice (Bakker & Gravemeijer, 2006);
- index of inequality (Kung, 1993; Littlewood & Polya, 1988);
- research evidence and foundation for knowledge construction (Cooper, Levin, & Campbell, 2009);
- workplace or recreational tool for decision-making or self-monitoring (Masinglia, 1994; Naresh, 2008; Nasir, 2000);
- human development point of origin for abstract knowledge (Dewar & Xu, 2011; Xu & Garcia, 2008);
- vehicle for organizational sense-making (Weick, 1995);
- socially symbolic interaction of professionals (Morrissette, 2010);
• materialization for blending concepts (Hutchins, 2005);
• surreal myth (Webb, 2009);
• cultural symbol (Geertz, 1966);
• logical or moral proposition as the warrant for assertion (Dewey, 1941);
• form of organizational information (Easton, 1957);
• numeric text (Rotman, 2000);
• statement of student opportunity (Germain v. Ontario (Minister of Education), 2004);
• expression of scientific rationality (Popper, 1959);
• policy straightjacket (Graham & Neu, 2004); and/or
• instrument of state power (Foucault, 1986).

2.6 Significance of the Study

This study defines an average as a numeric sign for managers to guide their own actions, and for management to guide others’ practices. Understanding how leaders conceive of that sign, and reason with it has the potential to improve schooling in several ways. First, it promises to lend sophistication to existing notions of “data use” which have heretofore framed the role of statistics in educational administration. Scholars in cognitive psychology and behavioural economics have now rejected the excessively rational assumptions of expected utility theory (von Neumann & Morgenstern, 1947): adults often do not make choices based on anticipated utility as expressed in numbers. Data utilization theories cloud the role of reader experience, and overlook the fine shades of meaning that a reader may bring to an apparently transparent, symbolic notation with an asterisk and \( p \) value.

This study will also shed further light on our understandings of administrative agency, as school administrators’ ability to infer and act on the basis of descriptive statistics. Data utilization models sometimes presume that data follow a teleological path in cognition—from apprehension in reading, to information, to knowledge, to action
without wisdom (Ackoff, 1983, 1989)—and that statistical significance should be
governed by that mono-logic. However, the meanings that the reader makes with a report
may evolve in a circuitous or recursive manner. Moreover, data use models imply a
lockstep and linear path between reception and reaction, denigrating the school
administrator’s role to that of actor before the statistician-author as script writer. In short,
data utility models cast the school administrator as subservient to the author of data,
erasing agency and foreshortening interpretive processes with averages. What is missing
is a coherent theory of interpretation that will enable us to link perceptions of data
accessibility and interpretation through to purposeful and thoughtful action: this study
attempts to make a small contribution toward development of such a theory.

Depicting the reader of statistics exclusively as actor or consumer does little to
assist with the preparation of thoughtful instructional leaders and the cultivation of a
numerate adult learner. A key adult education issue is knowing how the innumerate
reader moves from a position where the average dominates the reader to one where the
reader exercises dominion over the number. From an organizational theoretic point of
view, we need to know more about how the reader makes sense of that data that swirls
within and around an organization—sometimes mischaracterized as a sense-making
entity (Weick, 1995) standing apart from its members—and how school principals'
interpretations intersect with those around them. From a performance management point
of view, little research is available which reveals how the leader as reader of statistics
ascertains strengths, weaknesses and areas for improvement in schools using the average
as a lens. From an evaluative perspective, we know little about how numeric values
intersect with professional or leadership values and purposes. And from a public
administrative perspective, the ability to interpret the statistical counts that often sit in the
middle of formal financial or government-generated accounts to create personal or narrative accounts, lies at the heart of improvement and accountability, as obverse sides of the same statistical coin.

Of course, policy-makers have an interest in averages as performance indicators, wanting to more clearly communicate trends. For example, at the federal level, a recent Statistics Canada report on international education indicators (2010) contains 43 tables or charts of data demonstrating student attainments, budget expenditures, graduation rates, annual earnings, and other measures of school system performance for the provinces and territories in Canada. A total of 796 Canadian and international averages are provided within these tables as points for national and international comparison, and thousands more if the comparison points are enlarged to encompass the provinces themselves. In no case does the report describe in detail how the average was computed, whether a participating country was included or excluded from the comparative average, or the assumptions that underlie the particular formula used. Quebec’s latest provincial education indicators report (Ministère de l’Éducation, du Loisir et du Sport, 2011) relies on averages for depicting pupil-teacher ratios, average annual expenditures and average international test scores for students. Bridging between the notions and notations of practical and statistical significance, and the micro-and macro-dynamics of data use (Moss, 2012) is thus a policy problem, as well as a practical communications issue.

Investigating the processes that school principals use when reading statistics may also yield insights into pan-Canadian challenges with numeracy. National studies (Statistics Canada, 2003, 2005, 2013) indicate millions of adults lack rudimentary quantitative competence; international comparisons reveal that difficulties with ratio concepts, in particular, are widespread. Numeracy as the analog of literacy refers to an individual's ability to understand and use mathematical information at school, at work,
and in everyday life; for example, handling money and budgets, using measurements for cooking, or reading a map. Numeracy is a foundational skill for individuals and is associated with better opportunities for life-long learning, as well as success at work (Hoyles, Noss, & Pozzi, 2010). Because principals are conversant with educational issues, can articulate learning difficulties and processes with insights unavailable to a lay audience, and are leaders in organizations, schools and communities, they are well positioned to provide insight into the challenges of creating a numerate population. Moreover, principals themselves are increasingly required to become more numerate in the context of their duties—to learn on the job rather than seek formal retraining in an institutional setting—which will be a challenge for the larger adult population. Although principals as skilled managers have diverse academic and social backgrounds, their numeracy skills may not match new workplace demands (Organisation for Economic Cooperation and Development, 2013).

The average is also important for professional researchers who want to write publications that are easy to understand. As a summative index, the average is widely used in not only statistics and psychology but across the social and natural sciences. Many research findings in journals are reported as means, and inferential statistics revolve around differences between means. In mathematics, the integral is a weighted sum, while in physics, the centre of gravity and the centre of mass are weighted means. In public and business administration, methods such as cost-benefit analyses or linear software programming use weighted means to evaluate alternative courses of action. A failure to specify between the mean and median, notwithstanding safety code requirements, may explain over-engineering costs by as much as 30 percent in civil engineering (Young, 1989). Neural science has modeled the neuron as a cell body that computes a weighted mean of inputs; computer science does the same with artificial
neural networks. Meteorologists and stock market analysts rely on weighted means when making predictions, and the media and polling agencies estimate proportions of people approving government policies using weighted averages.

Moreover, the average is often set as an educational policy target. In Saskatchewan in 2011-2012, for example, the Ministry of Education had a goal that “student achievement in mathematics, reading and science reach the Canadian standardized average” (Saskatchewan Ministry of Education, 2011, p. 6); it has shifted its focus now (Saskatchewan Ministry of Education, 2013) to close disparities in graduation rates between First Nations and Metis students and non-Aboriginal students by 50 percent before 2020 (Howe, 2012). Both objectives necessitate nuanced conceptions for the mean. School boards are responding, but we do not know how or indeed whether school administrators are. As one local example among many, the Regina Public School system (2012) depicts gaps in average high school marks between First Nations and the general high school population in its Continuous Improvement Framework reports without indicating either the computation or the authors of such averages. Nearly every school division now reports annually to its publics using normative comparisons of some type, as do postsecondary institutions. In short, we know little about the way that school administrators who work with either Aboriginal or non-Aboriginal students conceive of the average student, and interpret school board plans that feature averages.

2.7 Delimitation and Limitations

This study’s focal point is conceptions for measures of central tendency—their apprehension, comprehension, implication and expression from the point of view of the school administrator. Computations are considered only insofar as they reveal beliefs (Pollatsek, Lima, & Well, 1981). Terms such as “student” and “achievement” are
defined within Peirce’s theory, and deemed as ancillary concepts or perceptual lens through which the notions of an average, as a noun, may be qualified. For that reason, this study is qualitatively dominant. Thus, although an average is often considered in an adjectival or modifying sense, my concern is how school leaders consider the average as modifying its object, whether the student or his or her achievement or staff or school in the first phase of this investigation. In the second phase, my concern is strictly confined to how conceptions of the average may modify school administrators’ own beliefs about their approaches to problem-solving.

My focus is on the statistical result, not on the classroom or large-scale assessment processes that generate the statistic. Many educators uphold “authentic” principles and are familiar with how classroom grades are generated, but they do not have extensive expertise with the technology of testing; they may be prone to discussing the effects of testing rather than the implications of a test score. Of course, any study of consequences or effects implicitly draws in conceptions of causation. However, the primary zone of inquiry in this study is what the average “causes,” not what “causes” the average. In other words, my interest is the ways in which a conception of the average may express, a post erori, the causal agency of its reader. Therefore, the study carefully excludes attributions of causal forces prior to the conception, except through those frames that may alter respondents’ habitual beliefs.

Although this study largely adopts a constructivist approach, it does not delve too widely into social-constructivist approaches to mathematical cognition, nor offer conjectures about broader culture-cognition relationships. French-language ideas for measures of centre are not considered (Antoine, 1998; Barbut, 1991, 1998; Feldman, Lagneau, & Matalon, 1991). For reasons of manageability and practicality, the context is circumscribed as situational and relational in its first phase, contingent in the second
phase. This is not a study of “cognition in the wild” outside the principal’s office (Lave, 1988; Saxe, 1991), nor in the pristine laboratory of totally-controlled conditions, but within the quotidian venues of administrative work. Rather than opening up a broad line of investigation through an ethno-mathematical study, a pragmatic concern with economy in research has repeatedly led me back to the narrower and more immediate confines of those ”mathematics in the work place” studies that have been conducted elsewhere.

The first, qualitative phase of the study is thus limited by the extent of my immersion in a school setting, relying only on the testimony of school administrators, without extensive observation of or co-participation in actual leadership practice. I do not assume that multiple interviews with 10 Saskatchewan educators are generalizable, although their insights and experiences may be transferable. Nor can I suggest, given the limits in sampling frame, that the results in the succeeding quantitative phase will be generalizable in a statistical sense beyond the borders of Saskatchewan. My overall sampling was not completely randomized, although the assignment of participants within the experimental design was.

My aim, therefore, is to counterbalance strengths and weaknesses in these two stages of inquiry, trusting that mixing methods will reveal different yet complementary facets of an average. Neither phase of this inquiry documents actual decisions, but rather dispositions within decision-making processes when considering average student achievement. Insofar as a model of decision making in education is imposed, this study deals with school principals’ choices within cyclical practice rather than any formal or informal regime mandating a dichotomous choice, as prevails in cognitive psychology.

Above all, I am conscious of un-resolvable paradoxes in using qualitative means to illuminate mathematical matters. In phase 1 of my study, I rely largely on transcripts of interviews that are verbal in nature. At issue is whether linguistics mediates or
mangles mathematical understanding. I am aware that statistical understanding can and should be communicated statistically. However, Brown (2001) argues that mathematics becomes “a subject of hermeneutic understanding if the emphasis is placed on interpreting mathematical activity” (p. 49), which is also a linguistic activity. In other words, the meaning of an average may not be susceptible to direct hermeneutic consideration, but averaging as learning and interpretive process for generating mathematical meaning is susceptible to linguistic and logical expression. The difficulty with this approach revolves around the ontological status of matematized objects in themselves: it might imply that they have no meaning except insofar as we can express averages in words. An equally serious concern may arise when I use various types of means to analyze school principals’ conceptions and decision making with averages: sampling, analyses of covariance, and regression all unavoidably draw on notions of a mean. To some, this may seem like the dog chasing its tail in an ongoing tautological search for meaning. However, I do detail the means that were employed in my analyses.

This study thus deals only with school administrators’ gloss for an average insofar as their conceptions touch instructional leadership functions. I do not delve deeply into the many other facets of the administrative role—staff recruitment and professional development, financial management, facilities planning, and additional tasks that may invoke in some way, other conceptions of the average. This study revolves around administrators’ roles in relation to teaching, learning and assessment; in organizational theoretic terms, the study is delimited to withinputs in light of outputs, not inputs (Easton, 1957). The focus is strictly on Saskatchewan school administrators’ concept(ion)(s) of “the average” in relation to student achievement.
2.8 Recapitulation and Organization of the Dissertation

Drawing on abductions from various research traditions and policy statements in North American education, I provisionally hold that:

1. Managing the meaning of statistics is part of a school administrator’s role.
2. Researchers have not previously inquired about school administrators’ interpretations of elementary descriptive statistics in relation to student achievement.
3. Interpretation of a statistic precedes and overlaps with managerial action.
4. Closing gaps in student performance as represented by statistics is a key objective of many (inter) national, provincial, and local policy-makers. School administrators may have the same objective.
5. Our understanding of the “average” in the phrase “average student achievement” is important for closing educational gaps, and thus becomes a key issue in school administration.
6. Canadian school principals, in their legislated duties, have a crucial role in improving student performance, as mediated through the efforts of others. Educational administrative practice in many of its facets involves the interpretation and use of basic statistics such as the average.
7. Scholars hold different perspectives on the school principal’s role and how/which statistics have salience.
8. The way school principals conceive and calculate “average” student achievement in relation to consequences is a key question. The effects of an average framed in alternate ways for instructional planning is another key question for instructional leadership.
9. The statistical representation of an average may have both qualitative and quantitative properties.
10. A distinction may be drawn between the author of a statistic and the reader of that statistic. In parallel, distinctions may be drawn between statistical significance and practical significance.
11. An average as evidence can be used to make, inform, legitimate or promulgate a decision.

If Chapter 2 of this dissertation projects some of my aspirations for an average, Chapter 3 describes the English-language research about adults’ interpretation and use of elementary statistics—that is, some of the aspirations projected on the average to date. Chapter 4 revisits some of the ideas of pragmatism’s progenitor, specifically in relation to
practitioners’ meaning-making with statistical signs. Chapter 5 outlines the
phenomenological and cognitive psychological methods adopted in this study to explore
school administrators’ conceptions of the average. Chapter 6 describes 10 Saskatchewan
school administrators’ conceptions of average student achievement within leadership
practice, as perspectives. Chapter 7 traces the influence of perception on 210
Saskatchewan school administrators’ pre-dispositions or beliefs when reading averages.
In this way, I have taken an inventory of the prospects for averages within school
administrators’ decision making. The eighth but not conclusive chapter draws together
both qualitative and quantitative findings toward a holistic but provisional view of school
leaders’ interpretation of averages.
Chapter 3: Review of Literature on Averages

North American and European research within several scholarly disciplines is reviewed in this chapter, and critiqued in relation to alternate conceptions of the average. Research assumptions and disputes about the meaning of an average are considered in terms of the primacy of mathematical forms, sociological norms, cultural formations, functional uses, and cognitive failures in interpretation.

3.1 Mathematical Forms of the Average

Mathematicians generally assume the meaning for an average inheres to its form through deductive reasoning. In tandem, statisticians sometimes presume that meaning arises within a comparison of two averages through inductive application. Both recognize an average as a portmanteau term, enveloping the mean, the median, the mode, and the midrange. All these conceptions of the average involve notions of variation, distribution, proportion, and probability. Measurement specialists will emphasize two types of means: a sampling mean, which describes the representativeness of the sample from a larger population, and the measurement mean, which refers to the event or activity that is the object of investigation within that universe. In all cases, statisticians who deal with averages will specify their conception through an algorithm before proceeding to generate a sign or notation of central tendency. They presume that anyone with an elementary education has a rudimentary understanding of that simple algorithm. Moreover, statisticians as authors presume that when they compare two means, and generate a simple inferential statistic with an asterisk and the letter p, its meaning or significance may be transmitted externally to another party, the reader.
In contrast to the statistician’s conception of the mean, median or mode, mathematics educators have thus far identified several alternate conceptions that students and teachers as readers appear to hold for measures of central tendency (Garcia & Garrett, 2006). The *average as typicality* revolves around the representativeness of individual scores, but remains vague in overlooking variation (Garfield & Ben-Zvi, 2007; Mokros & Russell, 1991, 1995). In the *fair share conception*, frequently modelled in elementary pedagogy, the average becomes a point of reasonableness for the allocation of resources. Other researchers have found that readers conceive the average as *signal-or-signature in the midst of noise*, as a way of summarizing and depicting a disorderly set of observations (Groth, 2005; Konold & Pollatsek, 2002, 2004). Some readers may hold a *flow conception* for an average as the velocity or rate by which an object moves in relation to another object (Thompson, 1994). As a *point estimation*, the average may also be understood as an intermediate or best value from a controlled series of observations (Bakker & Gravemeijer, 2006; Batanero, Cobo & Diaz, 2003; Garcia, Cruz & Garrett, 2006).

For others, the average becomes a *centre-of-balance*, where the distribution of points in a data set finds an equilibrium point (Barr, 1980; Mevarech, 1983). Those hewing more closely to algorithmic ideas (Cortina, 2002; Cortina, Saldanha & Thompson, 1999) see the *average as a multiplicative* or procedural add-and-divide, a conception which most elementary teachers seem to hold. This latter type of quantitative reasoning is in itself complex: the quantity itself must first be conceived as an entity for which a person constructs a schema when thinking about situations, and about objects which have measurable attributes—as well as those for which zero has been assigned. Hence, the mean is rooted in images of quantification itself: the reader must understand which attributes are measured and measurable, and must have tacit knowledge about the
performance of the object. Readers thereafter draw on procedural knowledge, and make adjustments to understand the ratio of contributions to contributors. In a recent review of the statistic, Garfield and Ben-Zvi (2007) concluded that “What has been striking over 25 years of research is the difficulty encountered by students of all ages in understanding concepts related to distribution, center, and variability” (p. 386).

To illuminate distinctions in mathematical form and meaning, some researchers have traced the phenomenon of an average historically (Bakker & Gravemeijer, 2006). Some didactic phenomenologists make the dubious claim that ontogeny will recapitulate phylogeny—that individuals serially reiterate each historical conception within a human development trajectory. The mean, as is the case for mathematics overall, originated either independently within or in commerce among different cultures, using different symbolic notations—Chinese, Indo European, perhaps Japanese. Axiomatic proof was the strength of the ancient Greek mathematician; Chinese mathematicians focused on place value, decimal device computation (the abacas), algorithm development, and algebra—the weaknesses of their Greek counterparts. It is said that the ancient Greeks studied 10 different forms of the mean: linguistic and cartographic cognates came to include mediate (in the middle of opposing parties), medium (communicative channel), mitten (wrapped about the hands and capable of manipulation), meridian (a navigational line), and Mediterranean (the middle of the world, at least to the ancients).

Three historical developments illustrate the increasing sophistication brought to both computing and applying the mean. The first key application of the mean was in maritime calculations during the high tide of mercantilism in 15th century England. Etymologically-speaking in English, the word average denoted "damage sustained at sea," thus tracing its lineage to Lloyds of London and maritime insurance. Averaging was originally a way of mitigating and managing risk by equalizing a burden (Hardy,
Littlewood, & Pólya, 1988). The type of calculation used in adjusting general averages conflated the concept of "average" with "arithmetic mean" for insurance and taxation purposes, whereas in French, the “moyenne” has consistently denoted “the middle.”

A second key development was the nineteenth-century application of the mean to social phenomena and not just natural events, thus providing a foundation for the emerging social sciences. Although Laplace and Bernouilli had foreshadowed the application of probability to the measurement of uncertainty in human phenomena, Quetelet, Poisson, Galton, and Pearson were all influential in adopting and refining the average in relation to “the average man”—for estimating birth and mortality rates, determining the available heights for military recruitment, examining crime rates, estimating literacy levels, and even uncovering the “laws of history” (Porter, 1986; Stigler, 1986, 1992) The method of least squares, binomial distributions, and fitting distributions to data all demonstrate new applications of the mean and median in late nineteenth-century social science.

The third historical development was the application of the mean to the intellect. Wundt, Fechner, Ebbinghaus, and others in late nineteenth-century Europe were proponents of what was called “psychophysics,” as studies of perception (Goodwin, 2008; Stigler, 1992). However, Americans like Terman, Spearman, Thorndike, Thurstone, and Otis pioneered many early principles of psychometrics (Jones & Thissen, 2007)—originally for efficient role differentiation when raising an army during World War I—that became popular across American education as universal, compulsory secondary schools were implemented in the 1920s. Harold Foght in 1918 brought such Progressive ideas to Saskatchewan, when commissioned by the Government to recommend reforms to the provincial school system. The American specialist in rural education found that the “system of examinations in use is a questionable norm of the
average pupil's scholarship, ability, maturity and fitness for advancement” (p. 8). To buttress his recommendations for the creation of a compulsory high school education system in the Province, with consolidated administrative structures, and a revised and reduced examination system based on American developments in psychometrics, Foght made extensive use of surveys which drew on computations of the mean (Foght, 1918).

In brief, mathematicians and statisticians can point to many forms of an average which have deep roots in Western history and which have been available to Saskatchewan educators over the past century. Mathematics educators have identified alternate conceptions for these measures of middle held by students and teachers. No research has been conducted into First Nations’ traditional concepts of centre, either within or outside counting practices. Nor, surprisingly, has there been any investigation into educators’ calculations of weighted averages, which sit at the heart of nearly all secondary school grades, although an initial study among high school students reveals much computational confusion (Gattuso & Mary, 1998). We know little about school principals’ conceptions of the average, and the way they think about and with various measures of central tendency for describing student achievement.

3.2 Sociological Forms for a Norm

Of course, for many people, an average is implicitly a summative representation for a norm. A norm can have both statistical and sociological meaning, and many people thus associate the average with individual and societal characteristics or attributes. The sociological inculcation of norms was a central goal for public education well before Saskatchewan entered Confederation. The term “normal school” was derived from the French École Normale, an institution that provided inculcation in the “norms” of school instruction for teachers in the wake of the French Revolution, seeking to replace the church order with secular public schools for both boys and girls. The term was
transplanted to teacher training institutions in Ontario by Egerton Ryerson in the mid-1800s, migrating westward thereafter. When Saskatchewan’s first Normal School was established for training teachers in the early 1890s, the explicit mission of the school was to ensure that teachers were prepared to inculcate Canadian norms in school children. At that time, the norm was to be personified by the teacher. Concurrent with the establishment of Normal Schools in Saskatchewan and Alberta, the territorial and eventually the provincial Ministry of Education instituted Departmental examinations at the Grade 8 and 9 levels in the early 1890s, to regulate entrance into high schools and to qualify teachers for teaching therein. The first Board of Examiners was formed by a Protestant minister and a priest in the 1880s and, by 1898, exam scoring sessions were conducted by the North West Teachers’ Association (Black, 1913). Since the Province’s beginnings, transcripts of marks at the high school level have been based on percentages, not an alpha scale of A through to F, nor the British four point classification system (personal communication, Gerry Craswell, 2012). Thus, average student achievement in Saskatchewan has for long been associated with the percentage scale and setting educational norms.

During the 1980s and 1990s, the norm-setting function in Saskatchewan migrated away from teacher preparation and into the written curriculum as a statement of desirable student performance objectives. By the late 1990s, however, large-scale assessments were introduced in elementary and high schools to set norms for student achievement. Other Canadian provinces have witnessed this same transformation, whether in Ontario in the transportation of norm-setting functions from the original Ontario Normal School to the Education Quality and Accountability Office, or in British Columbia from the Vancouver Normal School to Victoria’s Ministry of Education. With the move to large-scale testing in most jurisdictions, the norm may therefore have become more
clearly associated with statistical forms rather than embodied in educators’ behaviour or instructional functions.

However, sociologists have long wrestled with the idea of normative beliefs, behaviour, rules and concepts, concluding that confusion in our definitions of a norm reflects the conflation of various types of norms. The absence of a stable typology leads to social scientists’ continual creation of *ad hoc* kinds of norms (Kahneman & Miller, 1986), prompting a protracted debate within cognitive psychology itself (Gigerenzer, 1996, 2001; Kahneman & Tversky, 1996; Vranas, 2000, 2001) about the meaning of a norm. Even statistical experts’ meanings for the norm have evolved over time, yielding a family of two-parameter curves such as a Poisson or Fourier or Cauchy distribution rather than a single Gaussian curve under the central limit theorem (Stahl, 2006). In fact, the concept of the norm can refer to regularities in behaviour, to a particular shared frame of reference, or to the existence of a social obligation or pressure. This results in much conceptual confusion. Norms have become synonymous with, alternatively: conventions, cultural mores, legislative decrees, behavioural standards, or cognitive development. Gibbs’ (1965) taxonomy delineates 19 different types of norms, but within sociology, there remains (1) a lack of agreement in generic definitions of a norm, (2) no adequate classification scheme for distinguishing types of norms, and (3) no consistent distinction between attributes of norms that are true by definition and those that are contingent on the social group being described.

Likewise, economists have recently attempted to clarify the meaning of a norm: Interis’ (2011) simplified typology emphasizes that economists regularly fall prey to the naturalistic fallacy, failing to distinguish between the descriptive “is” and injunctive “ought” when talking about normative behaviour. At the same time, his typology distinguishes between those normative definitions that involve social and self-sanctions,
and those that don’t. In fact, no definition suggests that a norm denotes strict uniformity in behaviour. Most sociological and legal conceptions of a norm are ambiguous, leaving many questions concerning the character of norms unanswered. For example, must a norm govern, control, or just describe conduct? Stated otherwise, how much deviation is allowed before the standard is no longer a norm? Also, must norms be supported by, or otherwise consistent with, collective values, in the sense that most persons find them just and acceptable (Gibbs, 1965)?

Such a socioeconomic reformulation of the average from form to norm illustrates another paradigm shift underway in Canadian education, from a nearly exclusive focus on inputs and process to include consideration of outcomes or results. The policy focal point is shifting from money spent to measurement of outcomes, from the contemplation of average-per-pupil expenditures to average student achievement. Where once the interpretation of financial data or numbers associated with allocating resources or students to programs was the predominate activity, the interpretation of student achievements has become important. Now school leaders must pay attention to outcomes, plan with results, interpret statistics, explain (away) outcomes, and (im)mobilize community interests in student learning. However, many educational administrators may not have a strong background in either measurement or statistics.

What links the mathematical average’s various forms with the sociologists’ norms is measurement, so the measurement specialists’ assessment forms become relevant when considering the meaning of an average. The terminology of averages is most frequently used in conjunction with norm-referenced measurement, which proceeds from different premises than criterion-referenced measurement. Whereas criterion-referenced
measurement revolves around measurement of properties such as student skill, knowledge or ability according to exogenously-determined “objective criteria,” norm-referenced measurement determines a student’s standing endogenously in relation to other examinees. That is, norm-referenced tests are explicitly designed to maximize the variability of scores, whereas criterion-referenced tests will likely generate a more homogenous distribution, and thus be less useful for rank-ordering individuals on a measured property. Norm-referenced tests are usually anchored at the middle or a mean group performance, whereas criterion referenced measures are anchored at the end points of a scale, at 0 or 100 percent for example.

Of course, purposes are paramount in test design. Norm-referenced tests are frequently used to make comparisons among examinees or to address “fixed quota” selection problems. On the other hand, criterion-referenced test scores are used to make descriptive statements of what students can do, to make instructional decisions, and to evaluate programs and their effectiveness. Examinees are judged on their own merits in selection situations, primarily in relation to an extrinsically-applied standard. Both forms of measurement are used to sort. Norm-referenced measures accomplish the sort by comparing the performances of individuals, whereas criterion tests sort in terms of mastery of a certain skill.

Not only is the mean a key concept in all measurement acts, it also enables the transmutation of scores between both forms of measurement. Criterion-referenced measures can be compared in terms of average proportions of populations achieving a criterion standard, such as whether a particular school has higher averages of students who pass as compared with another school. In turn, norm-referenced measurements can be translated into criterion measures by applying an extrinsic standard as a threshold to the normative scale to make a decision, such as saying that all Saskatchewan graduates.
must have an average high school mark of 75 percent in key subjects to meet a minimal
criterion for university entrance. In that sense, a criterion value threshold is a
superimposition, an overlay on a normative conception.

Such translatability, however, impinges on averages as evaluation points for
educators when they are considering a set of scores. In education, the statistical average
per se is frequently not a decision point, but rather serves largely as a point of reference.
Fifty percent is the usual criterion threshold for a pass-fail decision in school grading,
whereas 67 percent is often the anticipated average for a classroom set of scores, as a
point of reference. However, at the individual student level, an average below 50 percent
as a report card grade may be used for determining those who will move to the next grade
level. It might also be emphasized that any scale and not just a percentage or percentile
may be considered as an average. Whether a grade is registered as an alpha A to F; or as
a criterion 1 to 4; or as above/meets/below expectations; or as a chromatic red, yellow,
green; all may enfold a set of professional observations that supply information about
central tendency for consideration and interpretation.

In short, the various mathematical forms of the average carry much sociological
and educational freight, drawing in normative ideas about customary behaviour, cultural
mores, testing procedures, social sanctions, and statistical curves. School administrators’
reasoning with the average as a statistical representation for a norm has not heretofore
been investigated.

3.3 Formulae for Computing an Average

Of course, numeric forms issue from formulae. An average is both a
computational and conceptual act: for many American college students, knowledge of the
mean seems to begin and end with a computational algorithm. Researchers note
(Pollatsek, Lima, & Well, 1981) that understanding the concept of the average requires
several types of knowledge. Even simple problems require an operational knowledge of the averages’ “real world” applications, procedural knowledge in computation, and analog knowledge that involves visual and kinesthetic images of the mean as “middle” or “centre.” Moreover, different concepts for the arithmetic mean appear to arise even when the same problem is presented in different contexts. Depending on the school principal’s numeracy, context and purpose, then, she or he may draw on several meanings for central tendency as a representation for the "middle" value within a data set.

In doing so, administrators may associate the average with an underlying normal or Gaussian distribution, commonly known as the bell curve. The normal distribution is used to describe the general, or idealized, shape of graphs of normally-distributed data. The average “plays as significant a role in the social sciences as differentiation does in the natural sciences” (Stahl, 2006). The term central tendency is used in some fields of research to refer to what statisticians sometimes call "location." Other statistics, such as the standard deviation and the range, are called measures of variation and describe distribution or dispersion. Statisticians often offer the rule of thumb that “dispersion precedes location.” Many different descriptive statistics can be chosen as a measure of location beyond the arithmetic mean, the median, and the mode. The most popular is the arithmetic mean, but depending on the nature of the data, other types of central tendency may be more appropriate. For example, the median is used most often when the distribution of the values is skewed with a small number of extreme values, as seen with house prices, incomes, or very small classrooms.

In fact, school administrators may draw on many different formulae. Even Pythagoras distinguished between the arithmetic mean, the geometric mean, and the harmonic mean. For the arithmetic mean, the algorithm appears straightforward: dividing the contributions by the number of contributors. On the other hand, the geometric mean
of \( n \) non-negative numbers is obtained by multiplying them all together and then taking the \( n \)th root. It is often used for describing proportional growth or returns on investment; and may be more appropriate for comparing learning growth among classrooms. A harmonic mean may be more “intuitively” understandable, since averages track movements in time or space such as average speed for a number of fixed-distance trips, for example, school bus routes. A well-known issue is the non-equality of arithmetic, geometric, and harmonic means for any set of positive numbers, even if they are commensurable. That inequality has been formalized as \( AM \geq GM \geq HM \).

The computation of any mean becomes even more complicated when the author of a statistical average assigns weights to any value in the numerator of the arithmetic, geometric or harmonic mean. Little attention has been given to how people weight groups of judgements (Shah & Oppenheimer, 2011); for principals, this means the weight they would assign to a combination of language arts, mathematics, music, and French for example, to determine the students’ average performance. Moreover, calculations of an average may be subject to rounding of many types: convergent, Dutch, Gaussian, odd-even, or bankers’. We do know that the Saskatchewan Ministry of Education rounds marks at the Grade 12 level for high school graduation and credit purposes, but we do not know whether and how school administrators may round teachers’ classroom marks, particularly in promotion decisions.

Studies indicate that as many as one quarter of pre-service teachers confuse the mean and mode (Groth & Bergner, 2006; Leavy & O’Loughlin, 2006)—creating many challenges for teachers and educational leaders. The most frequently occurring number in a list, the mode, is not necessarily well-defined: the list \( (1, 2, 2, 3, 3, 5) \) has the two modes—2 and 3. The mode as measure of centre has the advantage that it can be used to represent typicality with non-numerical data (e.g., red cars are most frequent), while other
averages cannot. For an instructional leader, understanding the most frequently-occurring categorization of a grade for a student within a classroom or school has many implications.

Superficially, the median is less problematic. As the middle number of the group when numbers are ranked in order, consideration of the median can extend to the interquartile range, truncated means, and Winsorized means, where values are systematically removed from a computation of the “centre.” Moreover, some consider standard deviations as falling with the bounds of an average. Yet trimming or Winsorizing a data set to reduce dispersion raises many educational and indeed moral issues, particularly in a public forum or in relation to programs. Little is known about how administrators classify or exclude, either systematically or extemporaneously, those part-time, transient, special education, exchange, or newly registered students in their conceptions, calculations and representations of average student performance.

In short, when discussing an administrators’ average, we need to be precise about which measure of centre is being applied to which data. Technical reports are often posted for Ministry of Education large-scale assessment computations in other provinces, but are not in Saskatchewan. No research has been published to date about administrators’ proclivity to read such technical documents, how school principals’ read report cards, or whether they indeed read numeric information at all. A research void thus remains about administrators’ understanding of various averaging algorithms, and whether they are able to critically appraise or apply these formulae. How principals compute an average may reveal much about their conceptions and beliefs.

3.4 Cultural Formations Around an Average

Cognitive anthropology has affirmed that adults display visible and sometimes idiosyncratic practices with arithmetic concepts that have become customary within particular groups of people. Likewise, social psychologists point to social relations,
particularly accountability relationships, as influencing meaning. Although the mathematical practices of some professional groups have been studied for their mathematical understandings and behaviours, the practices of school administrators have not. School principals’ interpretations for an average may be bound up in workplace demands, accountability relationships, and expressions of power.

Any conception of an average, many cognitive anthropologists contend, will follow various cultural or societal formations. That is, the reader’s social and cultural contexts shape conceptions of statistics in daily practices that are situated in an environment, and performed differently for various audiences. Anthropologists argue that the reader’s practices with statistics are embedded in customs (often routines of a cyclical nature), which could explain difficulties in transporting averaging concepts across domains. Gravemeijer and associates have distinguished between four levels of “situation” for calculation. The first is a task level (in which interpretations and solutions depend on understanding of how to proceed computationally in a specific setting (most frequently on the job). Second is a referential level in which people’s conceptions refer to activity in instructional activities (most often in the classroom). A third, general level, features “models-for” that enable a focus on interpretations and solutions independently of situation-specific imagery. And the fourth is at a “formal level” which involves people reasoning with conventional symbolizations, which are no longer dependent on the supports offered in the previous three circumstances (Gravemeijer, Cobb, Bowers, & Whitenack, 2000, p. 243). The school principal may consider averages in all four of these types of situations.

Cognitive anthropologists often try to put a face on numbers. The perceived author of a statistic may be either imagined or actual: Nolen (1995) has demonstrated that when students know the name and personality of the author for a statistical textbook, there is more identification with the material. In contrast to developmental psychologists,
who are usually preoccupied with invisible mental mathematical operations in the working memory, cognitive anthropologists have focused on visible operations in the actual workplace. Alternatively described as everyday mathematics, street mathematics, ethno-mathematics, or out-of-school mathematics, their studies look closely at numbers for their practical significance. The arithmetic practices of dieters (Lave, 1988), carpet layers (Masinglia, 1994), candy sellers (Saxe, 1991), and South African carpenters (Millroy, 1992) have all been studied in depth. The anthropological field has encompassed the shopping plaza, the hospital ward, the basketball court, the engineer’s cubicle, and the investment banker’s office (Gainsburg, 2006; Hoyles, Noss, & Pozzi, 2001, 2010; Jurdak & Shahin, 1999; Nasir, 2000; Noss, Pozzi, & Hoyles, 1999). Such studies (Rogoff & Lave, 1984) implicitly refute formal computational schemes to favour informal routines.

The original aim of such studies was to deflate the contention that mathematical competences can only be acquired by attending (Western) schools, although international test results from the Orient and especially Asian economic indicators have cooled some of that ardour. In short, these studies “discover” mathematics in non-Western societies as well as in the academic hinterland of the marketplace and shop floor, to legitimate the peripheral participation of those excluded from elitist pursuits. Indeed, many anthropological studies—whether Saxe’s (1991) field work in Brazil, Lave’s (1988) in the Los Angeles shopping mall, or Naresh’s (2008) with bus drivers in India—are prompted by a rejection of dominant Western mathematical assumptions of linear rationality. A purely descriptive pursuit of mathematical activities in a natural setting, it is presumed, will reveal people’s practices that are just as effective as formal classroom mathematics.

However, after scrutinizing ethnographic studies closely, some cognitive psychologists counter that actual mathematical practices in natural settings are simple
arithmetic procedures performed as prodigious feats of mental computation in highly specific, routinized domains. The mental mathematics in many everyday situations typically involves small and round numbers; the most usual arithmetic practices are addition, subtraction, and multiplication (for example, doubling or trebling). Although some adults follow non-customary procedures, the mathematics itself is not practically different from that in an elementary or middle school North American classroom (Greiffenhagen & Sharrock, 2008). Therefore, what anthropological studies have revealed is that problem formulation changes with the context and purpose for doing the mathematics, not the underlying axioms and arithmetic operations themselves. The anthropologists’ concept of “emergent goals” when doing a mathematical calculation is important for all research fields: the reader’s purpose may be clarified during the computational process and is not necessarily clear before undertaking the calculation.

Custom as recurrent social practice is an explanation for acting without thinking much, the cognitive anthropologist believes, particularly the small group traditions that arise in the local context. For Nasir (2000), recreational and classroom practices with statistics become highly routinized in particular domains. For Hutchins (2005), the statistical traditions arise in the coordinated actions of people in their work, whether the Indonesian navigator or the Boeing pilot. Saxe’s (1991) work responds to what he perceives as the shortcomings of the theories of Piaget and Vygotsky in terms of their ability to explain how “the mathematical understandings that have emerged over the history of a cultural group become the child’s own, interwoven with the child’s purposive problem-solving activities” (p. 230). Tools enable a variety of conceptual blends (Hutchins, 2005) to be accomplished efficiently in specific domains as a form of distributed cognition, not social cognition. They permit the coordination of differentiated tasks uniformly among individuals across distances and across time. This means that any study of a school administrator’s averaging must remain attuned not only to the tools
used, but also to how the leader has mathematized an object in the past, looking for continuities and changes within present and past practices.

Both cultural anthropologists and social psychologists assume that significance resides in social formations, but the latter tend to locate it in relationships rather than in routinized behaviour. Generally speaking, social psychologists see statistics such as an average from the perspective of accountability relationships, and uphold universalist ideals in ways that anthropologists don’t. Although social psychologists have not yet specifically focused on adults’ meanings for particular statistics, social psychological precepts in accountability are relevant to administrative conceptions (Green, Visser, & Tetlock, 2000) of the average. For Lerner and Tetlock (1999, 2003), accountability is a universal feature of social interaction. Although universal, the specific norms to which people are held accountable vary from one culture and historical period to another. As a crucial psychological link between the individual decision maker and social systems, accountability arises in the need for norm enforcement. The average is a statistical form of social control. Social expectations constrain virtually any decision that humans make; and conversation has the effect of diluting expectations (Tetlock, Lerner, & Boettger, 1996). Thus, averages reflect and refract contradictory aspirations.

Social psychologists emphasize that such accountability expectations exist in daily relationships and are not necessarily formalized. In fact, when looking at accountability relations, they propose three alternative archetypes to the intuitive statistician that is upheld in cognitive psychology. Biases in thinking may stem not from the deleterious influence of intuition within the mind, but rather from social pressure. In Tetlock’s (2002) taxonomy, administrators may be pragmatic politicians trying to cope with accountability demands from key constituencies within and exterior to organizations. Principals may be characterized as principled theologians trying to protect organizational values from encroachment, making trade-offs to protect sacred values.
Alternately, in-school administrators may be stereotyped as prudent prosecutors trying to enforce social norms (Tetlock, 1999, 2000, 2002) with a statistic such as an average. These role archetypes create different empirical and normative boundaries on judgements of what is the average, which are now labelled as errors or biases within cognitive psychology. More fruitful research programs into judgement and choice, Tetlock and Lerner (2003) believe, would see statistical meaning as tied closely to organizational position, rather than to a learning trajectory or the risky gambles of an autonomous decision-maker.

Thus, in the social psychological literature, social roles will determine the reader’s response to a statistic. To cope with accountability pressures, research has shown that people use an acceptability heuristic (conformity in pursuit of approval), pre-emptive self-criticism, defensive self-bolstering, and decision avoidance (procrastination and buck passing) as strategies to avoid accountability (Lerner & Tetlock, 2003). The manager of meaning will therefore make compromises or concessions to discordant values within or outside the organization. An average may be negotiated or bargained to effect such accommodations. In a study of school administrators’ computations, the researcher may view a weighted average as computed and interpreted to affirm solidarity with the normative order in a community. The corollary assumption would be that the administrator’s meaning for an average will arise only in a dynamic conflict or tension with a teacher’s average classroom scores in local accountability arrangements, as part of an accountability relation with central office, or through compromise with parents for whom educators operate in loco parentis. Calculation of an average is thus both an arithmetic and political act: the average has not only constituents but also constituencies.

What we do not yet know is how a school administrator may collect, interpret, or generate descriptive statistics within each of these strategies and assumptions. Notwithstanding a large advocacy literature arguing against making test outcomes public,
there is a dearth of empirical evidence which describes the public’s influence on school level decision-making with accountability data. Making test information public cannot be considered as synonymous with media publicity, since parents will inevitably receive individual student results via report cards, and the general community often receives information about exemplary student achievement through local press coverage of school awards ceremonies (Cabus & Witte, 2012). Because the school principal is often characterized as responsible for student level outcomes—both in provincial statute and in leadership literature—public attention to school results could reasonably be hypothesized as affecting principals’ conceptions for the average (Møller, 2012).

The power of the media to amplify and distort people’s conceptions of the average in schooling has been much conjectured, but has not been rigorously investigated. Nevertheless, European studies have examined the intentional and unintentional effects of control mechanisms such as external testing programs in education (Elstad, 2009; Pawson, 2002; Rönnberg, Lindgren, & Segerholm 2012) and have paid particular attention to those programs that feature government inspection visits and public performance indicators following testing. Three general conclusions can be drawn from British and Scandinavian studies: first, no evidence suggests that school inspections have positive causal effects on the quality of schools. Second, meta-reviews (de Wolf & Janssens, 2007) into the effects of publicly releasing performance information at the school level are unambiguous: although principals and teachers believe that performance indicators are important, parents and pupils take very little notice of such indicators when choosing schools. And third, side effects on the behaviour of educators, such as “window dressing” (creating a falsely favourable impression) and other types of “gaming” (taking advantage of the testing situation to drill or exclude students) have been documented, thus implicitly upholding a social-psychological perspective.
3.5 **Formats for Presentation of an Average**

Formats for computation and presentation of an average both precede and may also indelibly shape an administrator’s calculation and interpretation (Okan, Garcia-Retamero, Cokely, & Maldonado, 2011). If the medium is the message (Tufte, 1983, 1997), then both technical skill with and the range of options available in software tools may influence interpretation. An average can be represented pictorially or numerically; spreadsheets, calculators, and graphing packages are widely available in schools. All calculations of the average require the choice of a particular algorithm, and therefore implicitly draw on a formula, whether that formula is coded in software syntax or presented as desktop options. We know little about who and how those software choices are made—whether autonomously by the school principal at her or his desk, in concert with an information specialist in central office, or not at all because it is predetermined in a commercial software package. At issue is whether software syntax rules govern or significantly shape interpretation.

A related issue is whether the data display which issues from software choice influences interpretation. The chosen format for presentation—whether bar chart or scatterplot with accompanying icons and scalable vectors—may not alone predetermine interpretive meaning for an average. When viewing the visual re-representation of averages, whether as maps, histograms or pie charts, readers retain considerable autonomy in the range of interpretations to be made from an arithmetic mean. How is this so? Viewers bring their own perspectives and world views to a presentation. Painters evoke proportions not by following an objective rule, but by appraising proportions in relation to the angle from which the observer sees. It is not perception but, more precisely, perspective that determines the meaning brought to a visual presentation. If we rotate the visual presentation of an average, we re-frame the picture; if we foreshorten the normative range or extend the historical line of an average, our
perspectives may change and thus the reader as viewer is not constrained in his or her interpretations or conceptions of what has been depicted.

Another difficulty with the claims of those who equate statistics with closed semantic systems is the conflation of mathematics or algebra with statistics. The mathematical sciences in themselves are closed semantic systems because they build on deductive proofs within an elaborated symbolic system. In contrast, statistical work is usually predicated on probability theory, which presumes that in the application of numbers to observed reality, there are varying degrees of probability involved when apprehending any phenomenon. Statisticians do not assert certainty through proofs, but rather describe degrees of chance (or belief) that can be calculated. In other words, our notions of probability—whether as chance or a degree of belief, deterministic or fatalistic, may influence an administrator’s reading of an average. Administrators’ notions of likelihood have not yet been studied, particularly for their interpretations of notations like $p$-values or confidence intervals, or notations of measurement and sampling error, which often accompany both graphic and tabular displays of the mean in achievement reports.

If an arithmetic average is conceived as a working tool (Hutchins, 2005), it will reflect a division of labour in the workplace. In other words, school administrators may have only a partial understanding of the student or achievement “objects” which are represented graphically or numerically or both. Studies of professional scientists have shown they have misconceptions of basic scientific concepts when interpreting the icons and axis titles within graphs because they are unfamiliar with the work of other scientists, or even that of assistants in their own laboratories and sites of field research sites. Roth concludes (2003), after drawing on a European semiotic, that the correct interpretation of graphed trends, rates and averages depends on understanding the entire production process in scientific inquiry. Whether a blip on a graph is a significant sign or simply a
spurious event involves deep familiarity with the domain being observed, the data collection devices, the representation-producing mechanisms, the expressive domain of graphics, the feature-enhancing techniques in software packages, and translation techniques between representation and reality. Likewise, school principals may not understand the meaning of an average without having a detailed understanding of how teachers have generated marks in the classroom, with all the attendant assessment procedures and curriculum weightings involved. Moreover, when looking at groups of natural scientists at work, Roth discerns “a mathematization of professional vision,” not only a mathematization of the object that is being investigated (pp. 262-263). Those qualities in the natural world have been transformed into quantities that can be discussed. The underlying issue is whether teachers, when calculating a statistical average for various audiences, and administrators, when interpreting the average in numeric or graphic form, are similarly “reconstructed.”

3.6 Functions of Statistics in Education

Many studies of administration imply that the arithmetic or sociological form of an average will follow its functional use. When looking at the functions of data in education— that is, the intertwined roles that statistics and administrators perform within schools and the education sector—we can identify at least four research traditions: Knowledge Transfer; Modal Technologies; Social Constructivist; and Enlightenment. Each draws on alternate assumptions about the nature of a statistic, its impact in schooling, the causal agency of the reader, and the relationship between reader of a statistic and author of that statistic. For the most part, scholars in educational administration have been preoccupied with functional “uses” for data as an undifferentiated mass (Anderson, Leithwood, & Strauss, 2010; Coburn & Talbert, 2006; Coburn & Turner, 2011; Earl & Katz, 2002), rather than with conceptions for particular types of statistics.
The Knowledge Transfer orientation is the oldest research tradition, deriving from the natural sciences; it considers any number as a lever. Dating back to the program evaluation literature of the 1970s (Leviton & Hughes, 1981; Weiss, 1979, 1980a, 1980b, 1980c), researchers in this tradition presume a linear sequence of events within a positivist view of causality: basic statistical analyses lead to applied research, then to development of alternatives, and conclude with direct application in decision making or practice (Bradley, 1998; Rich, 1997). The underlying supposition is that the researcher as statistical author first identifies an opportunity that may have relevance for educational policy or practice. Applied research is then conducted to define a set of findings for practical action—which could range from a curriculum guide, to a new program, to new budget software. In this sequence, the statistical average depicts some aspect of reality, the leader-as-reader elaborates and refines the findings through further interpretive understanding, and ultimately the new knowledge is applied in a variety of settings, many of which may be far different from those in which the original data were collected.

Knowledge Transfer is now often called “evidence-based decision-making” or “knowledge mobilization” (Levin, 2002). A central assumption is that averages compound. The sheer weight of accumulated evidence as knowledge, captured and communicated with statistics, should lever the school administrator or audience to change behaviour or take a different course of action. Within this orientation to the function of averages, inductive reasoning is often assumed to lead the reader from the statistic toward action. Thus, within a particular setting, a descriptive statistic such as an average is presumed to move in one direction—from author to user, causing a group of people to think or act in a certain fashion.

Scholars of Knowledge Transfer draw distinctions between ordinary knowledge derived from practical experience, which is sensitive to context, and such scientific or formal knowledge embodied in a statistical presentation, which they often presume is
context independent (Cousins & Leithwood, 1986; Lindblom & Cohen, 1979). The characteristics of the administrative reader-as-recipient of an average also become relevant in this research tradition. The information required for decision making or policy execution, and the time pressures facing the user of an average become pertinent. Discussion moves beyond turn-around-time between information collection and dissemination to considerations of the timing of the arrival of a set of numbers in the school leader’s hands within administrative cycles. Decision characteristics such as the decision’s impact, type, novelty, and policy significance are identified as variables, as well as the policy orientation of the reader, the organizational position of the decision-maker, and the existence of intra- and inter-organizational rivalries. In addition, the decision-maker’s organizational roles, his or her information processing style, experience, social characteristics, and attitudes toward numeric data and evaluation are salient. Pressures from competing sources of information are also influential (Cousins & Leithwood, 1986).

Within this approach to the function of statistics, the application of numeric data is frequently considered inherently good and non-use a deficiency. Use of data is often deemed to signify the triumph of science, reason, objective evidence, and dispassionate social scientific or accountancy rules over the forces of inertia, agency politics, innumeracy, and the self-interest of managers and professionals. The presumption is not so much that organizational members lack other sources of information, autonomous judgement, and multiple means for making decisions or meaning from events. Rather, it is that statistics such as an average should prevail over anecdote, interest, or ideology, these latter considered distortions in educational endeavour.

A second approach to the way statistics function in schooling may be called the Modal Technologies perspective, which derives from a European view of instrumentality as a modus operandi: descriptive statistics are a prop of those in power. The most
articulate exponent of statistics as numeric appendages of centralized authority has been Michel Foucault (1994). In his consideration of *gouvernementalité*, or the rationalities of administration, the process of manufacturing statistics is synonymous with the technological means for manufacturing consent and exercising power. From about the 18th century onwards, he claimed, governing a population or an economy has involved exercising power over knowledge and laws by modulating their processes and conditions. Statistics became one of the key modalities for production of the knowledge necessary to govern. Governments produce information as part of their calculated control and subordination of populations, optimistically expanding data collection while declaiming the ends of wealth, public order, virtue, and happiness. There can be no well-ordered political machinery or disciplined administration, from this perspective, without statistical knowledge of the state of the population. The numbering of persons, goods, activities, births, crimes, deaths, and much more provides the material upon which administrative calculation can operate (Foucault, 1975).

Several writers (Miller & Rose, 1990)—following from Foucault’s views on *gouvernementalité* (Burchell, Gordon, & Miller, 1991)—conceive of statistics as policy instruments, as “intellectual technologies” that render aspects of existence amenable to calculation and hence administration. They suggest that graphs and charts of data, and particularly conceptions of an average, are instruments that give legitimacy to the designs of those in positions of power. Borrowing and adapting Bruno Latour's (1987a) notion of “action at a distance” to statistics, these writers argue that such indirect mechanisms rely in crucial respects upon “expertise”: the social authority ascribed to particular agents and forms of judgement on the basis of their claims to possess specialized truths and rare powers. In so relying, the self-regulatory capacities of subject citizens are shaped and normalized. Survey and census data, large-scale assessment information, and economic and social indicators represent government in action. They are the material traces of
complex acts of counting, that render conceivable, problematizable, and controllable vast ranges of socioeconomic and psycho-social activity across vast territories (Hacking 1991).

Integral to the Modal Technologies view are the formation of centres of calculation (Latour, 1987b). For the most part, these centres are in government, but universities bear some responsibility as well. To exercise power over events and processes distant from oneself, it is necessary to turn them into data that can be mobilized and accumulated. Events must be recorded in standardized forms, and statistics gathered. But aggregation, comparison, compilation, and calculation must be done in a central locale. Through the development of complex relays of record-keeping and accumulation, new conduits of power are realised between those who wish to exercise power and those who are to be subject to that power (Hopwood, 1987). Attempts to control the lives of others through policy are accompanied by a veritable avalanche of printed numbers (Hacking, 1991). Statistics turn subjects into normalized objects, enabling the machinery of government to control from central hubs that calculate.

Context is a coercive force, because the descriptive statistic, as a medium in itself, frames and controls the environment within which the reader operates. For some writers, such as Stephen Ball (2004), the terror found within performance statistics is performativity—that numbers are part of a neo-liberal project to suborn educational professionals into a numeric straightjacket that emphasizes action without thought. An average therefore becomes the state’s vehicle for erasing agency and structuring a population (Graham & Neu, 2004; Miller & O’Leary, 1987; Miller & Rose, 1990; Rose, 1988, 1991). Following this line of thinking, then, both the principal and professional teacher are centres of calculation and extensions of the state, because they generate and disseminate student marks. However, we do not know whether or how power and the provenance of a statistic about average student outcomes may colour interpretation.
The Modal Technologies view of how a statistic functions affords greater latitude for the reader’s interpretation than the Knowledge Transfer framework. However, the reader is nevertheless presumed to become docile, even punished by the statistical author. The commodification of knowledge and the intrusion of managerialist and accountancy perspectives in liberal education are part of a top-down conception of indicators as descriptive statistics. Performance indicators depicted with averages are a device used by politicians and business elites to dissolve public educational institutions (Graham & Neu, 2004). Indeed, and in direct contrast to Knowledge Transfer adherents, many writers in the Modal Technologies orientation see numeric information as inherently malignant, because those who create or commission statistics are misguided or even malevolent in their intentions. The reader’s non-use or resistance can be a sign of organizational progress.

A third outlook on the functionality of averages might be called the Social Constructivist orientation. It adopts a generative view of causality rather than the sequential view of those in the Knowledge Transfer and Modal Technologies orientations: statistics are vehicles for creating community. This tradition imagines the reader as (re)constructing a descriptive statistic according to individual propensities and conceives of statistics as texts that open in dynamic relationship with others (Cousins & Earl, 1995; Cousins & Walker, 2000; Kaser & Halbert, 2009; Roth & Bowen, 2001). The reader is not just an empty vessel into which the statistician-author pours numbers, but rather an active agent in generating and manipulating statistics or searching for other information for his or her own purposes. The statistical average, whether as social science research or as performance data, enters the decision arena as part of an interactive search for knowledge. When administrators read statistics, they are engaged in making choices and seek data not only from statistical agencies, practitioners, planners,
journalists, clients, interest groups, and aides, but also generate their own averages and construct their own graphs.

Thus, from the Social Constructivist vantage point, statistics become more useful at the local level by creating a sense of ownership. The assumption is that direct involvement of the reader as sole or co-generator of a table or chart of averages will cause that reader to experience self-determination and ego involvement. He or she will better appreciate the process of collecting, analyzing, interpreting, and communicating with data, and thus have a greater personal investment in its ultimate use. Interaction processes create social networks and also secure the reader as leader rather than as receptacle of the statistician-author’s average. This provokes members of an organization to challenge the cultural perspectives, values, and basic premises of others in the organization, creating the conditions for fundamental change (Cousins & Walker, 2000). The statistics then transcend extra-organizational boundaries and have the potential to dissolve social borders.

To Social Constructivists, the statistician produces a number not for a single reader but with and for a community of readers as decision-makers (Roth & Bowen, 2001). The statistician knows that “he or she will be interpreted not according to his or her intentions but according to a complex strategy of interactions which also involves the readers, along with their competence in language as a social strategy” (Eco, 1990, p. 67). Accordingly, the emphasis in this school of thought is as much on decision-making processes as on the data itself. Data becomes useful in different ways at each stage or type of decision, with intra-organizationally designed and conducted needs assessments being deemed more useful than extrinsically generated data that does not engage organizational members (Alkin & Stecher, 1983). Generalization is not the aim. Educational statistics can be used through conversation and storytelling to either alter or affirm those relationships (Earl & Katz, 2002; Kaser & Halbert, 2009).
A fourth research tradition to data use has been called the Enlightenment view, which emphasizes the transitory nature of statistics (Best, 2001, 2008). Proponents of the Enlightenment view recoil against the prescriptive role of statistics in aligning social reality with rigid mathematical or social models. Commentators often position educational statistics within a broad policy arena rather than within the local community. For such writers (Bryck & Hermansen, 1993; Darling-Hammond, 1992; Weiss & Buculavas, 1980), the relationship between statistics and socio-educative problem-solving is more akin to illumination than engineering. They believe that the pursuit of a “production function” or a collection of linear variables in education is illusory and misguided. Statistics may be useful in that they identify and frame issues and catalyze new ideas. Data can signal, as an early warning system, to help define new issues, offer new conceptual frames in which to discuss educational questions, provide useful information for initial brainstorming of possible solutions, and inform the broader public.

In this view, statistics such as an average will rarely provide specific solutions for school improvement that can be directly enacted, in direct contrast to the claims of those in the Knowledge Transfer and Social Constructivist traditions. Rather, the descriptive statistic is pertinent for initial stages of policy development or problem definition; it sensitizes readers-as-policy-makers to new problems and may convert non-issues into issues, facilitating the identification of alternatives (Weiss, 1980a, 1980b). Numbers help to shift the parameters within which solutions are sought, rather than directly improving school effectiveness.

Those adopting the Enlightenment perspective do not assume that the reader of a descriptive statistic always has purposes compatible with those of its author. Nor is there a presumption of the authoritativeness of the findings or the authority of the data producer, as in the Modal Technologies view. Statistics that challenge prevailing truisms may work their way into the official consciousness and, with support from dissidents and
other interests, may overturn customary values and patterns of thought. Nor is it assumed that decision makers as readers seek out statistics when faced with an issue or that they are receptive to, or aware of, specific statistics. On the contrary, statistics and the accompanying commentary percolate through informed publics and come to frame the way in which readers think about socio-educational issues. Whether descriptive statistics are used at all depends on the distribution of interests within and between various factions, the ideological commitments of participating groups, and the nature of their previous information, which influences their receptivity to a numeric presentation (Alkin, 1990; Weiss, 1983).

From the Enlightenment perspective, statistical averages are a type of news, a form of continuing public education that enables policy makers to update their map of the social world (Weiss & Bucuvalas, 1980). The numericization of educational affairs at the provincial and local levels facilitates multiple overlapping discussions. In the Enlightenment orientation to data use, statistics are useful insofar as they enable their reader to use or re-purpose them to illuminate larger social issues. Less emphasis is placed on the original intent of the statistician as author, though the quality of the statistics themselves is scrutinized to ensure their durability in debate.

From the Enlightenment perspective, a statistical presentation may function as a completely open text. The interpreter simply “beats the text into a shape which will serve his own purpose” as Rorty notes (as cited in Colini, 1992, p. 56; Rorty, 1961). Yet the interpreter or user of the statistic is still constrained by the intentions of the statistician through the discourse rules under which the original statistic was created. In fact, many Enlightenment adherents see statistics as a form of coinage to be traded. Although the statistic still serves as an existential proof, its documentary appeal lies in its ability to persuade or justify in support of a negotiation for resources or influence. For example, a school administrator may take the statistical presentation into a budget meeting for
negotiation or exchange rival graphic displays and tables with the mass media or parents’ groups in the jostle for power. Usefulness is thus cast in terms of a statistic’s potential for reinforcement—to warrant a claim or to secure allies and to undercut foes.

Insofar as those adhering to Enlightenment orientations attribute any philosophic notion of causality to data use, they often cast statistics within a catalytic model: statistics such as an average enable ongoing public debate, if not convergent understanding; and a general (mis)understanding titrates out of an indeterminate mixture of swirling public or professional opinion around an issue represented in sundry ways by numbers. Administrators who read statistics may subsequently use or misuse them deliberately or by accident. The dynamic of creation and critique ultimately sifts useful from non-useful data. Thus, there is an ongoing societal evaluation of statistics according to their efficacy as multipurpose and mobile resources, not according to their efficacy for obtaining compliance, as in the Modal Technologies perspective, nor by their fidelity to the object they describe, as in the Knowledge Transfer point of view. Numbers (and reports, recommendations, and other documentation) are currency that is exchanged in public discussion as educational organizations and individuals continuously define and adjudicate the goals and emphases of a public education system.

Statistical use in the Enlightenment orientation has a comforting quality. Commentators seem to promise that without any special effort, some plurality of truths will eventually triumph and that bad data will ultimately be cast aside in a kind of Darwinian process of statistical evolution or survival of the strongest number. Yet the Enlightenment view has its deficiencies. When statistics traverse diverse socio-political networks through indirect and unguided channels, they lead to invalid as well as valid generalizations. The only procedures for sorting out shoddy and obsolete data are found within the discourse rules of the respective community of interpreters, rather than through use of a normatively understood set of principles. Sometimes, unexpected or sensational
research findings, however incomplete or inadequately supported by data, enjoy the limelight. The indirect diffusion of numbers is susceptible to oversimplification and distortion, leading to obfuscation as much as to enlightenment. Statistics can be considered cynically, therefore, as a public relations tool, as nothing more than “data smog.”

In brief, those scholars who study the functions of data in education emphasize the broad uses of data rather than the practitioner’s specific inferences made with particular statistics. Questions about the leader-as-reader’s perceptions of causation with data, the power associated with statistics, the capacity of school administrators to create community around an average, and the impact of policy or publicity on school leaders’ meanings for an average, remain unexplored.

3.7 Leadership Finesse in Sense-making

The school principal, as manager of statistical meaning, must not only make sense of a numeric representation of average, but also help others within and outside the school do so as well. A substantial body of research exists on sense-making—the organizational theorist’s synonym for significance—and how leaders help others interpret events when making decisions, usually through verbal interventions. The seminal figure in such research, Karl Weick (Daft & Weick 1984; Weick, 1995), upholds narrative rather than numbers as crucial to sense-making, and contends that the entire organization interprets events, not individual members. Although few scholars of organizational sense-making have dealt specifically with statistical notations (Boland, 1984, 1993), such research is pertinent to ideas about the average. Like psychologists, scholars of sense-making frequently start from the position of failure, but at the organizational rather than individual level (Maitlis & Sonenshein, 2010). Initially, sense-making research focused on the interpretation of crises, and the leadership dexterity involved in interpreting low probability, but highly traumatic events like fires or corporate collapse. However, over
the past decade, organizational theorists have refocused to events with high probability but ostensibly low consequences, such as in organizational change. Increasingly, researchers in the sense-making literature are focusing on the micro-processes that underlie organizational change (Reay, Golden-Biddle, & Germann, 2006; Wrzesniewski, Dutton, & Debebe, 2003), and the ways that managers shape the meaning of change for others.

Sense-making theorists underline that different individuals and groups are involved in strategic change processes, not just the titular leader. One key factor that differentiates these studies is whose sense-making is examined. Most organizational theory focuses on top management. For example, Gioia and associates’ research on strategic change emphasizes executive sense-making (Gioia & Chittipeddi, 1991; Gioia & Thomas, 1996; Gioia, Schultz, & Corley, 2000). These studies bring into focus how senior leaders not only make sense of the external environment to formulate change strategies, but equally importantly, how they influence others’ construction of meaning through what is called “sense-giving.” Nevertheless, the goal of most sense-making is to understand how senior managers’ interpretive systems impact changes in organizational processes and outcomes (Thomas, Clark, & Gioia, 1993).

While executives play a central role in efforts to make changes, researchers have recently broadened their focus to consider middle managers during change. For example, Balogun and her collaborators (Balogun, 2003; Balogun, Hope-Hailey, Gleadle, & Willmott, 2005; Balogun & Johnson, 2004, 2005) emphasize the pivotal role of middle managers in mediating the sense-making of top managers with subordinate employees. Middle managers’ sense-making often moves from espousing old ideas to endorsing new ideas. Top management provides general information about proposed change, while middle managers are left to construct their own meanings of it. Middle managers, therefore, play a crucial role in rationalizing changes with other employees. Of course,
this is a top-down view of change—change can also be bottom up: middle managers’ interpretations may shape those above. At issue in centralized versus decentralized change management is whose conceptions and measures of centre prevail.

Specifically, Beck and Plowman (2009) highlight how middle managers have proximity to the interpretations of both strategic and frontline staff when change is underway. More generally, this research shows middle managers mediate sense-making between executives and employees on the frontline to affect both cognition and action. Therefore, managerial finesse is required because of disparate interpretations for events. Beck and Plowman (2009) noted that middle managers contribute to:

- a dynamic process of interpretation in which multiple filters from throughout the organization help frame and enrich interpretations of rare and unusual events. The interpretation becomes more like a mosaic of multiple understandings. The dominant concerns of strategic managers and front-line managers are woven into this interpretation (p. 919).

In sum, theoreticians of sense-making suggest that both immediate organizational failure and longer-term organizational change may impact on a middle manager’s meaning for statistics. However, no research has been conducted which deals with a school principal’s efforts to manage the meaning of statistics within or outside the school as organization.

3.8 Failures in Cognition

Notwithstanding the pertinence of the sense-making literature to the management of meaning, the most direct research into conceptions of the average in school endeavour derives from various branches of psychology. The main concern is why people fail to calculate the average appropriately. Through studies of failure, we gain a better insight into the arithmetic form and its evolution over a developmental trajectory. Theories in developmental psychology emphasize, therefore, the role of experience. This approach has three major strengths with respect to administrators’ conceptions of the average.
First, because a school principal was initially or often continues to retain some teaching responsibilities, the pedagogic and educational dimensions of an average become important when considering the administrator’s mindset. Pedagogical concerns have inspired the most concentrated attention to what the “average” means for didactic ends (Batanero, Godino, Vallecillos, Green, & Holmes, 1994; Cai, 1998; Cortina, 2002; Goodchild, 1988; Konold & Pollatsek, 2004). Research to date, which goes directly to educators’ conceptions of the mean and median, shows that elementary school teachers’ understandings of the arithmetic mean and median are shallow (Groth & Bergner, 2006; Jacobbe, 2011; Jacobbe & Carvalho, 2011), with some ability to articulate a procedure as formula but notable inability to apply appropriate forms in diverse situations (Callingham 1997; Leavy & O’Loughlin, 2006).

Second, because developmental psychologists are concerned with the production of individual knowledge, we have a relatively thorough understanding of the complexities of an average from the point of view of experts (MacCullough, 2007; Marnich, 2008). The difficulties in translating this deceptively simple concept between the statistical and the mathematical realms revolve around (mis)understandings of variation and contradictions between fair-share and centre-of-balance conceptualizations.

The third strength of a developmental approach is that it considers the relationship between mathematical cognition and emotion or affect (Evans, 2000; Mukherjee, 2011). Theorists operating within a developmental model have explored math anxiety in depth, along with its cognates—statistical anxiety, assessment anxiety, and accountability anxiety. Each type of anxiety demonstrates different symptoms and requires different prescriptions. Researchers have made a valuable distinction in adults’ mathematical thinking between attitude and emotion, generally seeing emotion as impairing attitude. So it is reasonable to assume that affective elements may be brought to interpretations of average student achievement.
When considering the way that adults interpret statistics, developmental psychologists often emphasize nurture whereas cognitive scientists often emphasize nature. Cognitive psychologists believe that biological endowment along with encoding processes in the mind determines meaning. They seek to explain why a particular person’s interpretation of a statistic differs from the norm. The study of heuristics and biases, as this research is usually known, revolves around tracing adults’ reactions to uncertain or risky situations as represented in numeric problems (Kahneman, 2002a, 2011). A foundational assumption within this approach to reasoning with numbers has been a dual-process model of the mind (Sloman, 1996; Stanovich & West, 2000), with two quite different cognitive modes, System 1 (S1) and System 2 (S2). The two systems correspond approximately to the common-sense approaches for experiential thinking and logical-analytical thinking. The two systems differ in various ways, most notably in the accessibility dimension: how fast and how easily things come to mind. Many of the non-normative answers people give to statistical (as well as other) questions can be explained by the quick and automatic responses of S1 and the frequent failure of S2 to intervene in its role as critic of S1. Based on this dual-process theory, Kahneman, in his Nobel Prize lecture, set an agenda for cognitive research:

To understand judgment and choice we must study the determinants of high accessibility, the conditions under which S2 will override or correct S1, and the rules of these corrective operations. Much is known of each of the three questions, all of which are highly relevant to the teaching of statistical reasoning (Kahneman, 2002b, p. 716)

Like developmental psychologists, cognitive researchers usually depict the reader as information-processor, and have been particularly influential in describing the inferential shortcuts, base-rate neglect, and aversion to loss that occurs when people consider probabilistic problems (Kahneman & Tversky, 1979). The foundational questions that cognitive psychologists address revolve around the information-processing
proclivities of the “typical” adult, and how well the “typical” adult makes decisions in optimal or utility-maximizing situations when reading a statistic. By far, the most extensive work on mathematical or statistical cognition has grown out of Kahneman and Tversky’s seminal work with Prospect theory, developed in contradiction to economists’ notions of rationality within expected-utility theory (Kahneman, Slovic, & Tversky, 1982; Kahneman & Tversky, 1979, 1984; Tversky & Kahneman, 1992).

Until the seminal papers in Prospect theory appeared in prominent economic journals, rationality had been typically associated with expected utility theory. This latter theory postulates that (a) people act so as to maximize their utility and (b) a preference order allows the choice of the option offering the maximum utility. Prospect theoreticians challenged those assumptions, asserting that people apply value weights to their decision options which can skew their choices away from those which, objectively or probabilistically viewed, may be in the best interests of the decision maker. In contrast to rational choice economists, Prospect theoreticians hold to three core principles. The first principle, reference dependence, proposes that people judge outcomes relative to a comparison point that divides the possible outcomes into regions of gain and loss. The second, loss aversion, proposes that losses are more painful than the same-sized gains are attractive. The third, diminishing sensitivity, proposes that people become increasingly insensitive to changes in potential outcomes as these outcomes fall farther from the reference point. These assertions have been repeatedly demonstrated through framing studies and depicted in an S-shaped value curve: its shape is concave for gains and convex for losses.

Brainerd and Reyna (1990) have posited a rival Fuzzy Trace theory to explain adults’ quantitative reasoning, particularly in the realms of risk assessment and judgement. Fuzzy Trace theorists claim adults do not reason at all with numbers, relying primarily on qualitative simplifications that are stored in memory to be retrieved in
problem situations. If Prospect theory is the traditional approach, proposing a formal, logical, and computational model for the application of values, Fuzzy Trace theory is based in Fuzzy logic and, questionably, in the mathematics of Fuzzy Set theory, because its theorists argue that reasoning is largely non-numeric in nature.

Under Fuzzy Trace theory, memory plays a preponderate role in reasoning with numbers, not prospective actions. Two types of memorial bases are operative. The first, a verbatim representation, is a ”surface form” copy whose detail is registered more or less exactly in words, pictures, or the natural numbers. It is a mental representation of the stimulus, not the stimulus itself. The second and parallel memorial base is a woolly, symbolic, mental representation that captures essential meaning. This fuzzy “gist” registers an event in a mediate form (Brainerd & Reyna, 1990) that is simplified, most often as binary opposites in simple categories. The two types of memory operate in parallel, simultaneously, and separately. Fuzzy Trace theorists hold, implicitly, that we cannot reason our way to new memories. The parallel but simultaneous nature of memorial function must be emphasized: fuzzy traces are not extracts from verbatim memories.

Multiple studies demonstrate conclusively that people have difficulty evaluating formally equivalent probabilistic statements in statistical form through shifting baselines or misrepresentations under Prospect theory. So too have Fuzzy Trace theorists demonstrated that one source for a “cognitive illusion” may be denominator neglect, the proclivity for people to weigh or code only the information in the numerator, ignoring important information in the denominator when operating on gist. In Prospect theory, an information leak has been posited to occur when tasks are framed. The leakage hypothesis and its supporting experiments are cited with approval by Fuzzy Trace theorists, who presume they are evidence of a gist in operation. In doing so, they overlook Kahneman and Tversky’s explicit description of the simplification in the initial
editing phase, where information is deliberately leaked. Likewise, a recent study (Kühberger & Tanner, 2010) pitting the two rival theories in different task situations suffers from this disregard for the editing phase.

3.9 Framing and Interpretation

Cognitive psychologists have identified several persistent biases when adults make judgements that appear related to risk. When statistics are framed in alternate ways within randomized experiments, problems have been identified in the interpretation of negatively- and positively-framed information, in adults’ perceptions of ratio-concepts, in controlling the mind to identify and attenuate biased interpretations, and in characterizing the operations of memorial and prospective reasoning when interpreting various kinds of statistics. None of these issues in interpretation have yet been explored among school administrators when ascertaining the significance of a statistic.

Research on risky choice has produced much evidence for framing effects, which may be defined as the impact of bringing different lenses to identical information. The choices that individuals make differ when equivalent options are described differently (Kühberger, 1998; Levin, Schneider, & Gaeth, 1998) under different conditions. In choosing between statistically-represented gains, most adults prefer the riskless over the risky option, whereas in choosing between losses, a majority prefers the risky over the riskless option. A school administrator facing statistics of above or below average student achievement may be contemplating or calculating risk.

A central debate within cognitive psychology revolves around the respective mental operations of intuition in System 1 versus the logic of System 2 in statistical understanding. Intuition remains very imprecisely defined, varying from a form of fast and frugal decision heuristic, as perception, as accumulated experience, as a cognitive illusion or trick, as a form of tacit knowledge, or as a bridging term in academic rhetoric that enables scholars with widely-different suppositions to communicate without especial
clarity or consensus, except in apposition (Dane & Pratt, 2007, 2009; Sadler-Smith & Sparrow, 2008). Cognitive psychologists usually operationalize their particular definition for an intuitive-versus-rational response in binary tasks. However, the intuitive statistician may be largely mythical because the very act of conceptualization is an act of rationalization: our concepts of any phenomenon, including intuition, thereby involve some kind of thought, not just arithmetic or statistical operations. Recent efforts have been made to consolidate the proliferating number of heuristics as intuitive-experiential approaches which characterize the fast and frugal way the human mind processes massive amounts of information (Shah & Oppenheimer, 2008), in contrast to the laborious approaches of a logical-analytic mind.

From an educator’s and leadership perspective, a key finding has been the negativity bias: formally equivalent (descriptive) statistical statements as percentages will be considered as more valid if framed negatively (Hilbig 2012a, 2012b; Kuvass & Selart, 2004). For example, “one-third of couples have divorced” may be considered as more truthful than “two-thirds of couples have remained married.” For the educator, this implies that communicating student achievement as averages may be adjudged as more truthful if negatively cast, but perhaps leading to demoralization of staff and students, bolstering the claims of critics while engendering perceptions of more honest appraisals among parents or others outside the school. Initial studies (Hilbig, 2009, 2012a, 2012b) suggest this bias arises from experience in processing information, particularly “fluency” which may be another way of expressing the “accessibility” of snap judgements in System 1 workings of the mind.

Meta-analyses (Kühberger, 1998; Kühberger, Schulte-Mecklenbeck, & Perner, 2002) corroborate the generalizability of bidirectional framing effects among physicians, mediators, and medical patients about particular drugs, and taxpayers toward audits (Chang, Nichols, & Schultz, 1987). Expertise in and knowledge about aesthetic appraisal
does not seem to reduce the “power of positive thinking” (Dijkstra, Van der Pligt, & Van Kleef, 2012) among those schooled in the fine arts when adjudicating paintings and other artistic performances as representations. Nor does knowledge of statistics and current events (Hilbig, 2009) help adults better discern truth statements. Additional thought and effort do not seem to mitigate the impact of such psychophysical distortions (LeBoeuf & Shafir, 2003). Given that neither theory has been tested among school leaders, we can only assume that, largely, undergraduate student proclivities apply as well to educational administrators.

Many other cognitive tactics may explain the framing effect: numerical recalculation; narrative reconstruction; the emotional influence of regret and elation; moral trade-offs (Tetlock, 1999, 2000; Tetlock, Kristel, Elson, Green, & Lerner, 2000); best or minimal outcome orientations; damage control; or consideration of outcome durability, to name a few. Administrators may make choices in terms of audiences before whom they perform, or in terms of multiple reference points (Koop & Johnson, 2012), as argued by Fuzzy Trace theorists. In fact, Kahneman himself has acknowledged the possibility of diverse starting points when we reason or negotiate with statistics (1992), distinguishing between reference points, anchors, and norms as representations of ambivalence. Administrators’ conceptions of an average may therefore be considered in any one of these ways.

Moreover, the different causal schema that readers hold may also induce them to “code” a task differently and to consider the events or issues within a framed problem differently. Lawyers may attribute causal relations to ethical relationships; engineers may attribute causation to mechanical relationships; medical professionals to physiological relationships (Maule & Villejoubert, 2007); and teachers to social relationships. Administrators may read causal forces into relationships between leaders and followers. In fact, we know little about the cognitive styles of occupational groups,
whether women are more subject to framing illusions, or whether leaders are more or less prone to rely on intuition as gist. Scholars have speculated that adults’ susceptibility to framing effects varies with educational levels, but not according to intelligence or competence. Moreover, without specifying temporal dimensions, we cannot really determine under Fuzzy Trace theory whether an exhortation to think longer about a problem makes any difference to the decision outcome, although thinking harder does not appear to do so (LeBoeuf & Shafir, 2003).

Ultimately such research agendas into meaning-making, unlike those broad functional positions adopted for data use, provide more precise lenses for considering how statistics may be conceived when making decisions in organizational endeavour. The controlled conditions of a laboratory have offered insights into the intuitive adult. Statistics are not treated as an undifferentiated mass as they are in functional or organization theoretic orientations. Cognitive science experiments invariably proceed from the position that decisions have a normative character, and that we can discern failures in cognition by measuring departures from the norm that are demonstrated by responses to basic descriptive statistics. However, cognitive psychologists have not heretofore looked at changes in problem-solving dispositions when adults interpret equivalent mathematical forms of the statistical norm. Nor have their research designs directly considered the particular purposes of readers, or their actual reasoning when interpreting a statistic.

3.10 Recapitulation

From previous research, we may adduce that:

1. As a mathematical concept, the average comes in diverse forms in both statistics and in our ordinary or daily practices.

2. Ideas about normality have been central to many educational practices from the beginnings of universal, compulsory mass education. Conceptions of a sociological norm as represented by an average are subject to dispute.
Alternate social and cultural formations have adopted different arithmetic practices that are visible when solving problems. Likewise, school administrators may demonstrate distinct practices.

The format(s) of presentation for an average may be influential in determining the reader’s interpretation and meaning for the average.

The computational procedures involved in averaging may reflect diverse beliefs and purposes in education. Calculation is not only a statistical but also a political or policy act; it may also be an organizational or social act accomplished with others in the school and beyond.

The reading of any statistic in terms of its function in decision-making is not a morally pristine act, but rather involves trade-offs and negotiation of sometimes conflicting value systems both within and outside a school.

Any statistic may be perceived and considered within two different facets of the human mind—one as rational and analytic and the other as experiential and intuitive. Cognitive psychologists posit a duality in the mind’s operations but do not concur on whether and how the two facets of the mind interact.

The role that intuition plays in school administrators’ decision-making with numbers has not been studied in any depth.

The way in which an average is framed or depicted may influence the manner in which it is interpreted by a reader. Whether a statistic is framed positively or negatively may determine whether and how an administrator will act upon it.

Mathematicians, statisticians, measurement specialists, anthropologists, sociologists, organizational theorists, developmental and cognitive psychologists have brought varying perspectives to the meaning of a mean, locating its significance in differing forms, formats and formations.

What is missing is an integrative theory that would bind these discordant perspectives together in a way that enables school administrators to exercise instructional leadership.

Pragmatism has long been considered to furnish a philosophic and methodological basis for practice. Revisiting the original ideas of Charles Sanders Peirce, as the progenitor of North American pragmatism, may offer further insight into the meaning and interpretation of statistical phenomena. What then, was Peirce’s theory of interpretation?
For the purpose of this inquiry, a Sign may be defined as a Medium for the communication of a Form.

Charles Sanders Peirce

Chapter 4: Pragmatic and Semiotic Ideas of Charles Sanders Peirce

In this chapter, the pragmatic ideas of Charles Sanders Peirce are described. His semiotic theory of interpretation, which integrates views of cognition, causation and behaviour, is indistinguishable from a theory of representation.

4.1 Charles Peirce’s Life and Location in North American Scholarship

Over the past two decades, several cognitive psychologists and mathematics educators have turned to Charles Sanders Peirce’s theories for insights (Bakker & Hoffman, 2005; Kehle & Cunningham, 2000; Presmeg, 2006) and for research approaches (Tashakkori & Teddlie, 2010a, 2010b) to reveal learners’ mathematical cognition (Cunningham, 1998). Peirce’s ideas about statistics as signs are especially pertinent to a study of school administrators’ interpretations of an average because he is widely recognized as the progenitor of American pragmatism as a philosophy. His theory of signs is now considered the most comprehensive to date (Houser, 2010), and therefore germane to analyses of school principals’ statistical reasoning.

Peirce’s vocation was the study of methods of science and logic, or, as he called it, the methodeutic of inquiry. Moreover, Peirce devoted a large proportion of his career to thinking about and through mathematics. Although Peirce did not directly address the central problems in educational administration of power and organization, he was nevertheless interested in psychology and measurement as forms of cognition and scientific practice. Peirce’s focus was mind, and his interest was in explaining the logic with which humans minds symbolically construct their imaginative universes. Practical,
pluralistic, participatory and provisional are adjectives that have been used to characterize his ideas. His propositions about “purpose” and its relationship to action, as well as his ideas about the connections between “purpose” and “cause-and-effect” relationships are particularly relevant for an investigation into the way school administrators reason with averages as numeric indicators.

Through John Dewey, a student of Peirce’s, pragmatism is generally acknowledged as having had widespread influence in North American education, although few will recognize Peirce’s name or his contribution to Dewey’s thinking (Prawat, 1999, 2001). And through Peirce’s lifelong friendship with William James, now recognized as one of the founders of North American psychology, some of Peirce’s ideas have permeated studies of cognition. In recent years, scholars in cognitive psychology and mathematics education have rediscovered Peirce. His ideas are applicable not only for mathematics education (Cunningham, Schreiber, & Moss, 2005; Kehle & Cunningham, 2002; Semetsky, 2004), but also for public administration (Shields, 2008; Snider, 2011; Whetsell & Shields, 2011), and educational administration (Hoy, 1996). Prominent educational thinkers have started to critique the research of others from a Peircean perspective (see for example, Noddings in Paul, 2005).

A practicing scientist for over 30 years, Peirce (1839-1914) was the son of a renowned Harvard mathematician and, as a Boston Brahmin, from a socially, politically, and intellectually prominent family. A polyglot, a polymath, and a philosopher of logic, his formal education culminated with a chemistry degree summa cum laude from Harvard. He served on the staff of the United States Coast and Geodetic Survey which used pendulums to explore gravitational forces across North America. Peirce was also associated with the Harvard Observatory, from whence came the one book he published during his lifetime, *Photometric Researches*. Later, he lectured at Harvard on the early
history of modern science and logic, and at Johns Hopkins University. However, his Baltimore appointment was not renewed, in circumstances almost Shakespearean in their tragic qualities: an intemperate personality, erratic professional and personal behaviour, marital infidelity, academic disputes with colleagues, and fears of agnosticism in a socially conservative academic milieu all played a part in his exile from the academy (Brent, 1998).

Despite being exceptionally prolific—his life’s work extends to some 80,000 pages (100 library shelf-inches) of manuscript that are only now being systematized—Peirce toiled in relative obscurity for the last third of his life. His writing was largely neglected in the first half of the last century, likely because of the circumstances surrounding his dismissal from Johns Hopkins. Yet Peirce’s benefactor, William James, repeatedly acknowledged that all of his thinking could be traced back to Pierce’s 1878 writings. Peirce continued to correspond with and chide John Dewey until at least 1905. Another of Peirce’s graduate students, the economist Thorsten Veblen, shared many of Peirce’s attitudes and translated some of his ideas into institutional economics at the beginning of the 20th century (Dyer, 1986; Hall & Whybrow, 2008).

During the past two decades, Peirce’s star has risen as scholars have begun to recognize that his thinking foreshadows and clarifies their own. The anthropologist Clifford Geertz acknowledged that his views about the role of symbols in society were closest to those of Peirce, although he found Peirce too realistic for his taste (Michelsen, 2000). Karl Popper has suggested that many of his own prototypes and propositions about science derived from Peirce as “one of the greatest philosophers of all time” (Bird, 1989), and Richard Rorty (1961) noted that “Peirce’s thought envisaged and repudiated in advance, the stages in the development of empiricism which logical positivism represented . . .” (pp. 97-98). More pointedly, Louise Rosenblatt (1998), whose reader
response theories have imbued language arts curricula across North America, acknowledged the influence of Peirce and has spoken at meetings of the Charles Sanders Peirce Society which still meets annually and publishes papers quarterly. Lately, Peirce’s ideas have become more widely known through a Pulitzer-Prize winning study of pragmatic thinkers’ influence in early 20th century American psychology, education, and law (Menand, 2001).

Peirce’s original ideas about statistical meaning have been obscured by his academic ostracism and by the gloss put on his propositions by others (Behrens, 2005; Green, 2007; Leary, 2009). Scholars have often superimposed their own concepts and terms on his, ignoring his presuppositions, or have truncated his ideas to the point of distortion2. Furthermore, his obsession with inventing new but now antiquated terms has impeded contemporary understanding. Peirce has been considered “a philosopher’s philosopher,” so it sometimes remains a challenge to bring his abstract ideas to earth. Moreover, Peirce was prolix and abrasive: a “very snarl of twine” is how he compared his personality to that of William James. Hundreds of pages of his writing wait to be systematically archived. His later writings were often drafted on the reverse side of his earlier manuscripts, making it difficult to trace evolutions in his thought. Editors and scholars have sometimes ignored his capitalization, considering it as unimportant3, and excised key passages. Accurately recovering Peirce’s theory thus necessitates returning to his original writings.

---

2 In a 1905 journal article in The Monist, Pierce noted that pragmatism had become so debased that he would henceforth label his own ideas as pragmaticism, thinking the term “ugly enough to be safe from kidnappers” (p. 166).

3 Regrettably, Peirce himself was inconsistent. In this study, a Sign refers to a representation, Final Cause refers to an idealized end state or purpose in the long run, and Reality refers to a fully realized end state in the long term. A non-capitalized sign refers to Peirce’s complete triadic view of the sign, final causation refers to pursuit of more immediate effects, and reality refers to the habitual beliefs currently held by a community.
What then, we might ask, were the central propositions in Peirce’s theory of interpretation, how do they compare with those of other scholars who study statistical reasoning in practice, and in what ways can Peirce’s thinking inform an inquiry into the meanings that school administrators have for “average student achievement?”

4.2 Pragmatism

4.21 Continuity and anti-Cartesianism.

Peirce’s pragmatism is a praxis philosophy because he considered the central feature of mind to be purpose, and purpose is always related to action. Adopting Alexander Bain’s credo that belief is “that which a man would be prepared to act upon” (CP5. 12), Peirce wove this assertion with many ideas derived from Kant and Scotus into a comprehensive doctrine of cognition that found its distillation in the pragmatic maxim:

Consider what effects which might conceivably have practical bearings, we conceive the object of our conception to have. Then our conception of these effects is the whole of our conception of the object (CP5. 402).

Peirce’s aim was to erase the boundary between thought and action. For Peirce, pragmatism was largely a method for clarifying ideas, not a philosophy. We must therefore carefully differentiate Peirce’s pragmatism from that of William James and John Dewey—where Peirce’s logic is replaced with psychology in the first instance, and with social action in the second. Dewey’s own forays into logic as natural history called forth Peirce’s critique of intellectual licentiousness and a “debauch of loose reasoning” (CP8. 240).

To render Peirce’s ideas clearly, I conform in style to that which most Peirce scholars use to cite his writings, but use APA in the reference list. His Collected Papers (CP), his Chronological Writings (W), his writings in the New Elements of Mathematics (NEM), his Unpublished Manuscripts (MS), his essential papers (EP) and his Cambridge lectures on Reasoning and the Logic of Things (RLT) will be cited throughout this study to indicate volume and paragraph or page number, rather than repetitively citing his name in the main text.
It is important to note that in formulating the pragmatic maxim, Peirce was not implying that an idea or concept should be evaluated on the basis of its practical value to the life of a community—this is a frequent misreading of what Peirce meant by pragmatism. When framing a conception's meaning in terms of possible effects or practical implications, Peirce emphasized that, since a conception is general, its meaning—that is, its intellectual purport—equates to the implications of its acceptance for general practice, not to a specific set of actual consequences or results themselves. He disagreed with William James’s popularization of pragmatism as “truth’s cash value,” or doing what is expedient (Houser, 2011). “Peirce wrote as a logician and James as a humanist.” Dewey (1984, p. 361) recalled as a former student of Peirce. Instead, Peirce reiterated that because humans live in communities defined by shared belief, pragmatism implied that the “clear” meaning of an idea held by an individual will be shaped in interaction with the beliefs—or habits of thought—held by members of the community as an indefinite organization. After all, he reasoned, “we perceive what we are adjusted for interpreting . . .” (CP5. 185). A conception’s clarified meaning points toward its conceivable verifications: actual outcomes are not meanings but individual “upshots”, to use Peirce’s terminology.

Like some later 20th-century researchers in the environmental sciences (Lovelock, 2009), Peirce maintained that all phenomena are interconnected or “continuous,” through continua; all philosophical or scientific endeavour should be guided by this principle, which he named *synechism*: “I do not merely use it subjectively as a way of looking at things but objectively put it forward to account for all interaction between mind and body, mind and mind, body and body” (MS 949). These three areas of interaction must be considered in turn for their pertinence to reasoning from, and through, statistics.

Peirce makes three assumptions about the reader of statistics and the reader’s stance that are not always shared by contemporary researchers. First, in his conception of
the relationship between mind and body, Peirce’s synechism—his belief that the mind and the body are inseparable—challenges many long-standing assumptions held in North American and European cognitive psychology. Many postulate that mind and body are separate entities, in a dualism that is widely attributed to Rene Descartes or to theological doctrines that hold that the spirit or “mind” of every individual is unique and severable from the body to become eternal in the afterlife. For Peirce, in contrast, Cartesian dualism “performs its analyses with an axe, leaving as the ultimate elements, unrelated chunks of being . . .” (CP7. 565). He believed we should not make overly sharp dichotomies between thought and behaviour. Peirce did not hold that physical and psychical phenomena are entirely distinct as categories of substance: all phenomena are of one character: matter is effete mind (EP1. 293). Though some phenomena are more mental and spontaneous, others are more material and regular: all are alike in a mixture of freedom and constraint, which allows them to be teleological or purposive (CP7. 570). In Peirce’s line of thinking, therefore, averages are not only constructions of the mind but also of the physical world and thus capable of formulation and expression in multitudinous ways. In other words, the mathematical or the statistical and the human form interact. This is embodied cognition, an area of research interest of late in mathematics if not in statistics (Calvo & Gomila, 2008; Lakoff & Núñez, 2000; Streeck, 2009). Peirce was adamant that such ideas will be expressed by the body. Excessively logical thinking which connects ideas to the brain, or as inhering to a soul, is misguided since:

. . . the idea does not belong to soul: it is the soul that belongs to the idea. The soul does for the idea what the cellulose does for the Beauty of the rose; that is to say, it affords it opportunity (CP1. 211).

Second, Peirce denied not only mind-body dualism, but also the notion that human minds are entirely separate from each other. Mind-body dualism and the insularity of human minds were ideas embraced by many of Peirce’s contemporaries,
most crucially by William James whose assumptions came to be held in much of American psychology. In James’s writing, human consciousness is conceived as being personal, thus leading to social pluralism. A scholar of Peirce, Michael Colapietro (1989) has noted this difference between James and Peirce: “. . . for the former, the most fundamental feature of personal consciousness is the irreducible fact of privacy whereas, for the latter, it is the ubiquitous possibility of communication” (p.78). Thus, Peirce’s theory of interpretation provides for a systematic explanation of individual and social cognition as minds interact. Peirce’s concept of a community as a “collective mind” undergirds his semiotic theory of interpretation, and is thus intrinsically normative because the meaning for a statistic accretes in social interactions.

Finally, in his allusion to the synechistic interrelationship of “body and body,” Peirce was reflecting some of the propositions in G. W. F. Hegel’s objective idealism. Peirce considered mind and matter to be but different aspects of the same totality, which he acknowledged were similar in some respects to Hegel’s “absolute mind,” but about which he had important reservations. Moreover, Peirce was persistently critical of Hegelian idealism, “a pitiful, club-footed affair” (CP5. 40; EP2. 144) which was deprived of a mathematical foundation (EP1. 256). In not recognizing the outward clash of experience—the interplay of character with circumstance—Hegel’s categories had the “coherence of a dream” (W5. 237). One characteristic of Peirce’s thought is his tendency to invest all ideas with biological properties of animate life and evolution. “Indeed, my opinion is only Darwinism analyzed, generalized and brought into Ontology,” Peirce once stated (EP1. 222). In doing so, he walked the tightrope between realism and idealism, although throughout his career he railed against nominalism, the philosophic tenet that definition exists only in words and not in things.
Peirce held that what we call matter is only an aspect of mind (and vice versa)—both are inextricably a part of the organic whole of the universe—part of this “absolute mind.” He presupposes that humans are organisms that have evolved through habit-taking to adapt to their environment; a human personality is a bundle of habits that is in constant communion through feelings and ideas with other humans. A personality “is some kind of coordination or connection of ideas” (CP6. 155). Peirce believed that just as a vast aggregation of cells constitutes every human being, human beings are in their turn “mere cells of the social organism” (CP1. 673). Society has a superordinate personality for in the case of the individual, consciousness is a kind of “public spirit among the nerve cells” (CP1. 354).

Peirce’s concept of collective mind derives from his conception of meaning. Peirce recognized that since members of a community share common struggles and questions, the meaning of their actions and thoughts must be predicated on the collective “habits of mind” of the community. Thus, for members of a community—whether a school, or a town, or a larger social unit—“what a thing means is simply what habits it involves” (EP1. 131). Peirce maintained that for members of a social circle, shared meanings are dependent on and reciprocally create the habits of the circle. That is, an individual’s understanding of what is average arises within a community, but in turn, their conception of the average then shapes the community and the individual.

Peirce further noted that the conceptions of truth that people hold individually are largely dependent on the habitually-held conceptions of their communities; rationality is not bounded by social or ethnic or even linguistic categories, but rather by routines. A key distinction in his outlook is that between “what we do not typically doubt” (tacitly held, provisional truth as it is habitually conceived by community members) and absolute
Truth (the ultimate, final answers to which all inquirers into a matter will eventually give assent). The implication is that truth for an individual is in some measure determined by his or her “circle of society.” This truth is likely incomplete and relative:

Two things . . . are all important to assure oneself of and to remember. The first is that a person is not absolutely an individual. . . . When one reasons, it is [one’s] critical self that one is trying to persuade. . . . The second thing to remember is that the man’s circle of society (however widely or narrowly this phrase may be understood), is a sort of loosely compacted person, in some respects of higher rank than the person of an individual organism. It is these two things alone that render it possible for you . . . to distinguish between absolute Truth and what you do not doubt (CP5. 421).

4.22 Belief and doubt.

Peirce recognized that the survival of every individual organism depends on the development of habits that will satisfy its needs. He observed that all human beings develop general patterns of coping actions in response to their series of life struggles: the whole function of thought is to produce habits of action. Concomitantly, people (and indeed all living creatures) acquire “habits of mind” or beliefs that guide their actions. According to Peirce, belief is not a momentary mode of consciousness; it endures for some time, and mostly (at least) unconscious. Frequently, however, chance events intervene to interrupt the habituated, inveterate actions of all people, creating doubt.

All thought begins with the irritation of doubt, arising from momentary hesitation in action or indecision that must be appeased. “Feigned hesitancy, whether feigned for mere amusement or with a lofty purpose, plays a great part in the production of scientific inquiry” (EP1. 128). To appease doubt, beliefs establish rules for action. A belief is a controlled and contented habit of acting in ways that will produce desired results only if it does not lead to further doubts. We acquire new habits in order to minimize the unexpected and the unforeseen, to orient ourselves and to minimize wasted effort as we pursue our goals, “to find our bearings like a ship in the open sea, with no one on board who understands the rules of navigation” (CP5. 368).
What a thing means is, therefore, what habits it involves, either cognitive or behavioural. A habit approximates what cognitive psychologists now label as heuristics, and what cognitive anthropologists call visible mathematics. A habit depends on how it might lead us to act under both probable and improbable circumstances. What the habit is depends on when and how it causes action, the first depending on perceptual judgements derived from sensations, and the second depending on a purpose to produce some sensible result. In *How to Make Our Ideas Clear*, Peirce thus argued

\[\ldots\text{we come down to what is tangible and practical, as the root of every real distinction of thought no matter how subtile [sic] it may be; and there is no distinction in meaning so fine as to consist in anything but a possible difference of practice. (W3. 265)}\]

Nevertheless, Peirce’s conception of a habit remains problematic, within his theory of interpretation, because of its elasticity. Peirce’s idea of a habit may encompass an involuntary but recurrent reaction, a biorhythm, a routinized behaviour, an organizational cycle, or a social custom. That is to say, it is unbound or open-ended, which is a recurrent challenge in applying Peirce’s ideas—indeed many philosophers’ abstractions—in practice. Conversely, much of the current cognitive science research into heuristic modes of thinking, Peirce would likely have argued, is an attempt to parse out issues in habit, not intuition, while failing to take into account the research subjects’ own immediate or long-term purposes. In fact, Peirce’s notion of a habit is remarkably similar to Kahneman’s (2011) technical definition for a heuristic as a simple procedure that helps find adequate though often imperfect answers to difficult questions (p. 98).

Categorization and purpose are interrelated in Peirce’s pragmatism, and therefore especially important in a study of school principals’ conceptions of the average. Peirce asserted that how individuals classify depends on their purposes; those purposes determine the characteristics of the class of objects under consideration. For Peirce,
purposes are “operative desires,” always general in character until the element of will is 
exercised upon an individual object on an individual occasion. As an individual pursues 
ends, that individual’s desires become more specific. Desires are variable, vague or 
indeterminate, and have a longitudinal dimension because the ideal state of satisfying a 
purpose is often not reached, or because others’ desires are not the same as ours. Because 
our desires are multiple, we strike compromises among them in their pursuit. Thus, the 
objects pursued will always cluster about middling qualities; such clustering distributions 
characterize purposive classes (EP2. 115-123). Peirce was concerned with distributions 
and purposive classes because “chance” as the calculus of probabilities, along with habit 
taking and law, are foundational elements in his universe. Within any inquiry into school 
administrators’ conceptions of average student achievement, therefore, we must pay 
particular attention to principals’ ways of classifying educational experiences as 
revelatory of administrative purposes, and their notions of probability as suggestive of a 
propensity to act upon an average.

4.23  Perception and intuition.

For Peirce, the elements of every concept enter logical thought “at the gate of 
perception and make their exit at the gate of purposeful action, and whatever cannot show 
its passports at both those two gates is to be arrested as unauthorized by reason” (CP5. 
212). However, as noted above, reasoning is not exclusively an affair of the intellect but 
has emotional and kinesthetic dimensions as well. Peirce uses the term “conceptions” 
rather than the term “concepts:” conception is integral to the fluid act of interpreting and 
implies possibility or potentiality, whereas a concept is too static for describing Peirce’s 
approach. In his view, “the function of conceptions is to reduce the manifold of sensuous 
impressions to unity” (CP1. 545)—to conceive is to simplify.
Unlike many 21st century researchers, Peirce rejected the claim that intuition is a human faculty. Peirce argued that there can be no power of intuition in the sense of cognition unconditioned by inference (EP1. 11-27). Our awareness of an internal world is by hypothetical inference from external facts. He also argued that there is no absolutely first cognition in a cognitive process; such a process has its beginning but can always be analyzed into finer cognitive stages. That which we call introspection does not give privileged access to knowledge about the mind—the self is a concept that is derived from our interaction with the external world, not the converse. At the same time, he persistently held that pragmatism and epistemology in general could not be derived from principles of psychology understood as a special science: actual reasoning patterns are too different from how we should think according to psychological models. That is a position to which some cognitive psychologists are now gravitating (Gigerenzer, 2008) in studying heuristics and decision-making. In his *Illustrations of the Logic of Science* series (EP1. 109-209), Peirce formulated both pragmatism and principles of statistics as aspects of a scientific method in general, as the best way to attain an accurate picture of the world.

Today, many cognitive scientists conceive of intuition within the controlled context of judgement. Peirce had a different perspective. He viewed intuition as informed judgement within the context of discovery. The largely pre-conscious processes involved in generating hunches are different from the conscious processes required to test them; thus, Peirce upheld the classical distinction between the context of discovery and the context of justification (Bowers, Regehr, Balthazard, & Parker, 1990). We might surmise that Peirce would see many current cognitive science experiments as studies of accountability effects, rather than as inquiries into intuition within cognition.
4.24 Induction and abduction.

According to Peirce, an inference is “the conscious and controlled adoption of a belief as a consequence of other knowledge” (CP2. 442; CP2. 144; CP5. 109). The inferred conclusion is true because in any analogous case, an analogous conclusion would be true (CP5. 130). The purpose of an inference is “to find out, from the consideration of matters already known, something else that we had not known before” (MS628. 4), thus to increase our grasp of truth. The references to “as a consequence of” and “because” in these passages indicate a premise-conclusion relation, perhaps suggesting causal significance since inference “produces or “creates” a belief in the mind of the reasoner (CP2.148). A belief, in turn, is described as “holding for true,” or “any kind of holding for true or acceptance of a representation” (NEM4. 39-40). To say we really believe in the truth of any proposition is to say no more than we have a controlling disposition to behave as if it were true (MS652. 15). The term “hypothesis” in Peirce’s lexicon, and perhaps the central idea in his philosophy, is a “proposition believed in because its consequences agree with experience” (CP5. 419); an hypothesis carries no suggestion of uncertainty or of something to be superseded, in Peirce’s use of the term (CP5. 420).

Three broad and distinct reasoning patterns may be traced in relation to hypothetic inference, Peirce insisted throughout the latter half of his career. Deduction proceeds from rule and case to result, as the formula of volition; induction proceeds from case and result to rule, as the formula for the formation of habits or general conceptions, a process which psychologically and logically depends on the repetition of instances or sensations. And abduction proceeds from rule and result to cases, as in the syllogism:

Rule: All the beans from this bag are white.
Result: These beans are white.
\[ \therefore \text{Case: These beans are from this bag.} \]
Whereas deduction explicates, and induction evaluates, only abduction (with its attendant hypotheses) “advances humans forward over the bridge over the chasm between the ultimate goal of science and man’s first wanderings in a primeval forest” (CP6. 475).

Peirce’s prototype for abductive reasoning was:

The surprising fact, C, is observed;
But if A were true, C would be a matter of course,
Hence, there is reason to suspect that A is true.

As a process, abductive reasoning generates a hypothesis as noun that explains our observation. To abduce therefore involves determinations of what is sufficient (or nearly sufficient) for A to be inferred from C, but not necessary. There are infinite possible explanations for any of the events or activities we observe, but we are inclined to abduce a single explanation (or a few explanations) for them in the hope that we can better orient ourselves in our surroundings and eliminate some possibilities. However, Peirce did not remain convinced that a single logical form covers all abduction, and he recognized a continuum of possibilities between pure induction and clear abduction (Goudge, 1969).

Simplification and economy are what call for the “leap” of abduction in thought. Toward the end of his career, he came to believe that a guess was one form of abduction (Kapitan, 1992):

The abductive suggestion comes to us like a flash. It is an act of insight, although of extremely fallible insight. It is true that the different elements of the hypothesis were in our minds before; but it is the idea of putting together what we had never dreamed of putting together which flashes the new suggestion before our contemplation. (CP5. 181)

Peirce maintained that humans must have some sort of an instinct that guides our guesses (CP2. 753; CP5. 172-4); they are not just chance insights. In abductive reasoning, unlike in deductive reasoning, the premises do not guarantee the conclusion. Abductive reasoning can be understood as inference to the best explanation, and formally equivalent to the logical fallacy of affirming the consequent, *post hoc ergo propter hoc*. 
In fact, Peirce equated pragmatism with the logic of abduction itself: it unfences our thought, allows for diversity, and prevents lock step interpretations from an average. However, Peirce counseled that such patterns in reasoning are much more than toys in Musement (contemplating or thinking about an object): “the Player should bear in mind that the higher weapons in the arsenal of thought are not play-things, but edge tools” (EP2. 437).

4.25 Diagrams and deductive reasoning.

Graphs and diagrams are central vehicles for conceiving or inferring from mathematical and statistical ideas in Peirce’s cosmos, and thus in practice as a habit. One hallmark of Peirce’s thought is his proclivity to infuse mathematical ideas with metaphysical properties. Metaphysics, Peirce claimed in his Guess at the Riddle of Existence, “is an imitation of geometry” (CP1. 354) and his synechism was a direct upshot of his belief in mathematical continua.

All our thinking is performed upon signs of some kind or other, either imagined or actually perceived. The best thinking, especially on mathematical subjects, is done by experimenting in the imagination upon a diagram or other scheme, and it facilitates the thought to have it before one’s eyes. (NEM1. 122)

Insofar as Peirce dealt specifically with the average, it was a probabilistic expression serving as a working hypothesis. In his Doctrine of Chances, he noted that newspaper editors, for want of precision, sometimes described the average man in terms of fractions: the average family consisting of 3.8 members bears no relation to reality. Until the turn of the century, he held a frequentist notion of probability, but had referred to a Bayesian logarithm for ascertaining degrees of belief, which he called the weight of evidence (Peirce, 1967). In Peirce’s view, the simple arithmetic mean expressed his view that chance was operative in nature. Chance embraces not only observations but arguments. The simple arithmetic mean reflects the chance that our arguments about
something are tenable. Average student achievement therefore expresses the likelihood that an individual or group would demonstrate particular qualities under specific conditions.

Peirce devoted enormous amounts of intellectual energy to formalizing and explicating his thought in what he called existential graphs, a painstakingly constructed set of diagrams to illustrate his reasoning toward a proof of pragmatism. For Peirce, the mathematized world is an imaginary universe, a virtual reality wherein one does imagined actions as reasoned bets or hypothetical statements. Seen in this way, both mathematical and statistical thinking are tentative propositions of conditional possibility or probability. All forms of scientific reasoning, he was sure, could be expressed in algebraic notation, in logic and in diagrammatic representation through his existential graphs. But he was never interested in mathematics as an end in itself; his motive for studying the algebra of logic was to discover with accuracy the primary forms of reasoning.

Diagrammatic reasoning in general and graphic logic in particular are integral to creating new scientific knowledge, according to Peirce, and thus expressive of an epistemology. His conception of a diagram was broad, perhaps excessively so: he identified three kinds—optical, tactical \( \textit{sic.} \) tactile?, and acoustic (respectively, these are represented by, for examples, blueprints, simulacra and sheet music) (EP2. 502)—that could be used as warrants for belief in a conclusion. He preferred the optical and believed that diagrams as systems of notation are required for an individual or group to comprehend any assertion. The reader of a diagram proceeds in three general stages, beginning with what Peirce called, at various points in his career, an abduction, retroduction, hypothesis, conjecture, or possible explanation which potentially explains some surprising phenomenon or experience that “breaks in upon a habit of expectation”
(CP2. 441). It is a spontaneous but plausible explanation and, once the inquirer has “the bit between his teeth” (EP2. 441), involves logically reasoning from the consequent to the antecedent.

In the second stage, that hypothesis is tested logically through deduction. This involves simplifying the problem by creating a diagram that strips away all extraneous elements and displays only the skeleton of the problem. “Only one kind of concrete clothing is permitted in a diagram—namely, such as, whether from habit or from the constitution of the mind, has become so familiar that it decidedly aids in tracing the consequences of the hypothesis” (Peirce & Moore, 2010, p. 19). At this point, the diagram maker-cum-reader follows two different procedures: first, she or he frames a pure hypothesis stripped of all features that do not concern the drawing of consequents from the conjecture. This she or he does without inquiring whether the hypothesis agrees with the actual facts or not. Second, s/he draws necessary or deductive consequents from that hypothesis. After explicating the hypothesis in diagrammatic form, and demonstrating its corollarial consequences, the inquirer proceeds to the third stage.

The third stage is inductive reasoning. Through creating subsidiary diagrams, she or he ascertains new relationships. Inductive reasoning involves determining the extent to which those new relationships accord with experience through classification, probations and judgements of the whole diagram. The inquirer may make subsidiary diagrams as modifications of the original to test and identify new relationships. Gradual inductive probations may be either qualitative or quantitative, and the inquirer will make a new estimate of the proportion of truth in the hypothesis, using measurements, counts, or statistics.

Therefore, for Peirce, the usefulness of a statistical notation was not primarily as a record of observations, nor as a metric, but rather as a device for introspection, for
conducting thought experiments. For Peirce, all reasoning could be diagrammatically represented. Such expressions have the same logical characteristics as the corresponding mental representations. A graph is a simple rendering of the statistician-cum-graph maker’s own mental representations, which reflect habits or interpretations of a problem or situation from which she or he can weigh the consequences or intents. The graphic representation is a spare depiction of the author’s experiences in light of her or his desires or purposes, enabling the author to externalize and thus contemplate the consequents for potential action. When interpreted by a reader (which may involve switching the role from that of author), the representation mediates between the author’s and the reader’s own past experiences and current purposes toward potential action.

Peirce tends to see a statistical presentation per se not as a communicative vehicle for measurements within a social setting, nor as a documentary record of a series of scientific observations for posterity, but rather as a device for self-clarification. Diagramming or statistical reading is a dialogue between the inquirer and his or her beliefs, rather than a way of establishing a direct correspondence with reality. The graph is at once a mental representation of thought patterns and cognitions, a platform upon which the imagination may be exercised, and a mediating device between the observationally real and the mathematically potential. In my interpretation of Peirce, the average—whether depicted in graphic, verbal or statistical form—thus becomes an experimenting tool which may serve nearly simultaneously as a device for concretizing abduction, a vehicle for deduction, or an instrument for induction from empirically-observed facts to cantilever the inquirer forward.

4.26 Scientific inquiry and statistical cognition.

Peirce’s notions of scientific inquiry were broad, and aptly captured in the German notion of Wissenschaf (Burch, 2013): we are all engaged in a (self-controlled)
quest for understanding; there is no rigid line between theory and practice, between teaching and learning, nor between statistical and practical significance. Peirce’s view of science therefore embraced both what we now characterize as qualitative and quantitative methodologies, spanning both phenomenology and the controlled experimentation now considered synonymous with the natural sciences. Although Peirce did not directly address questions of power and purpose in administration, he was explicit about the “methods” by which beliefs may be fixated in a community—that is, to avoid doubt (CP5. 358-387). His four methods follow.

The *method of tenacity* is used by those who have an instinctive dislike of indecisiveness. However, such persons eventually discover that they cannot hold their ground in practice, because the social impulse compels interaction with others, eventually shaking their confidence. Peirce’s understanding of this proclivity underscored his response to William James’ paper entitled “The Will to Believe”. Even though James had dedicated the paper to Peirce, Peirce countered with the title “The Will to Learn” (CP5. 583) because tenacity disregards evidence and arguments put forth within a community of inquirers.

According to Peirce, the *method of authority* was the primary means for sustaining theological and political doctrines. He opined that this method has governed the majority of mankind, from civilization’s beginnings. However, there are always individuals who have a wider social awareness than others, and who will doubt the “authoritative views” sustained by elites, and challenge these views.

The *a priori* method, on the other hand, appeals to an existing metaphysical system upheld within a community. It is usually held by people who have suspended inquiry because the fundamental propositions of such systems of thought are so comfortable. Nevertheless, the primary shortcoming of this method is that it cannot
consistently be verified through experience. Since each method had demonstrable shortcomings, Peirce concluded that the scientific method (that is, hypothesis testing by experiment) is the only satisfactory method by which beliefs may be settled to satisfy doubt. Peirce expressed what he called “the fundamental hypothesis of science” in the following way:

There are Real things whose characters are entirely independent of our opinions about them; those Reals affect our senses according to regular laws, and though our sensations are as different as are our relations to the objects, yet, by taking advantage of the laws of perception, we can ascertain by reasoning how things really and truly are; and any man, if he have sufficient experience and he reason about it will be led to the one True conclusion. The new conception here involved is that of Reality. (EP1. 138-139)

In his view, the scientific method of settling belief is “far more wholesome” because it allows “integrity of belief”—that is, the development of a conception of reality that is more well-integrating with the true Reality.

Since human conceptions are not necessarily revelatory of final Truth or Reality, Peirce concluded, it follows that all conceptions are best held provisionally. He noted that non-scientists misapprehend that little has actually been explained by science.

... notwithstanding all that has been discovered since Newton’s time, his saying that we are little children picking up pretty pebbles on the beach while the whole ocean lies before us unexplored remains substantially true as ever, and will do so though we shovel up the pebbles by steam shovels and carry them off in cartloads. (CP1. 117)

Recognizing that the results of inquiry may yield useful information, but are inevitably partial and humanly constructed, Peirce adopted a position he called fallibilism, acknowledging that any and all conceptions of Reality will inevitably be incomplete. He proclaimed, “The scientific spirit requires a man to be at all times ready to dump his whole carload of beliefs, the moment experience is against them” (CP1. 55).

He saw this as consistent with synechism, for “fallibilism is the doctrine that our
knowledge is never absolute, but always swims as it were, in a continuum of uncertainty and indeterminacy. Now the doctrine of continuity is that all things so swim in continua” (CP1. 171).

Not surprisingly then, Peirce was deeply concerned with scales and measurement principles (Short, 2008): his proposal to internationally define the metre using the frequency of light wavelengths was adopted and followed until the 1980s (Peirce, 1878, 1879). With his father, he computed probabilities using pairwise techniques for a court case involving forgery (Meier & Zabell, 1980). Peirce was also among the first to conduct controlled experiments with random samples in his investigations of telepathy, using a calculus of probabilities (Peirce & Jastrow, 1885). The concept of a normative distribution has been attributed to Peirce, although his “invention” was probably shared among a variety of statisticians such as Karl Pearson with whose work he was familiar (NEM3. 676). Yet Peirce warned against the threat to scientific progress of placing statistical science under externally-imposed social and political goals; he was nonplused by Pearson's pronouncement that the goal of science is to "strengthen social stability" as a form of Social Darwinism (CP8. 133). He also argued against the intrusion of provisional and tentative scientific conclusions, such as those of the higher historical criticism into practical affairs such as religion (CP1. 617-620).

A thorough and systematic study of Peirce’s contributions to and location within the evolution of the sciences remains to be undertaken and is beyond my purposes here. What is clear is that Peirce’s pragmatism is part-and-parcel of his very general conception of scientific inquiry and statistics, on one hand, and of a more phenomenological approach to the possibilities of investigation with documents as texts, on the other. Perhaps the most pertinent point to make is that Peirce’s thinking about statistics predates the significance testing of Fisher or the calculation of effect sizes by
Cohen (1988), and that he published contributions to correlation coefficients and theories of regression well before Galton. For Peirce, the significance of a statistic derives not from the endogenous comparison of two means, but rather from conceived effects or consequents as formulated within his theory of sign relationships.

Before turning to Peirce’s notions of significance, we might deductively summarize the central ideas in his pragmatism for school administration:

1. To ascertain the meaning of any intellectual conception we should consider what practical effects might result from the truth of that conception. The sum of these effects constitutes the entire meaning of any conception.

2. The human mind is an integrated biological and social organism within which ideas or conceptions are ensouled in a syncretic and continuous manner.

3. Belief and its privation, doubt, are central to understanding peoples’ habitual reasoning patterns and behaviours. Thinking and action are closely connected because what we think predisposes us to certain actions.

4. Conceptions are ways of reducing and simplifying our perceptual judgements of sensations. The ways we classify objects are directly related to our purposes.

5. We can conceive of an average as a sign of conditional possibility and as a sign of conditional probability.

6. Diagrams or graphics or notations of an average are a central pragmatic tool for reasoning abductively, deductively and inductively about reality.

7. People may hold beliefs about average student achievement tenaciously, as an authoritative statement, as an a priori truth, or as an object for further self-disciplined inquiry.

8. The average is a statistical formulation and application deriving from mathematics as pure hypothesis. Therefore, an average is an applied hypothesis whose meaning can be ascertained by holistically looking at its potential consequents and conditional effects.

9. Even though Peirce conceived of the average sign as a hypothesis or conjecture about the likelihood of events, other scholars and practitioners may not do so, out of habit.

4.3 Semiotics

4.3.1 Phenomenology and psychology.

Peirce’s mature philosophy revolved around his science of signs. In many ways, Peirce sought to do for cognition through an architectonic, what Mendeleyev had
accomplished for chemistry through the periodic table in the mid-1800s. Like the
chemical elements, the signs which operate in cognition have valences in Peirce’s
formulation. These valences derive from his sweeping, triadic categorization of all
phenomena. He never succeeded in clearly formulating the three fundamental categories,
drafting and redrafting them in successive versions of a never-published book. Peirce
identified three unique elements of consciousness: the realms of (1) pure quality or
sensory feeling, (2) brute fact or physical otherness, and (3) thought or synthetic
consciousness—as being the ultimate categories into which everything that is presented
to the human mind could be assigned. In 20th-century literary terms, they might be
called “sensibilities,” as illustrated by the different outlooks of the artist, the art dealer,
and the art historian on a single painting as a representation (CP1. 43).

Eventually, Peirce adopted the rudimentary labels of Firstness, Secondness and
Thirdness, respectively for his three fundamental categories, acknowledging the
limitations of language. These three categories are phenomenological rather than
psychological in nature. They are not categories of thought, but are so intangible and so
broad that they may be regarded “rather as moods or tones of thought, than as definite
notions” (CP1. 355). Firstness is the beginning, chance, freshness, evanescence,
originality, spontaneity, and freedom. It embraces the quality of feeling, and is entirely
separate from all conception or reference to anything else. Secondness is that which is
determining, cause and effect, terminating, ending, necessitating, correlating, or reacting.
Its facets include: otherness, relation, compulsion, effect, result, occurrence, negation,
and reality. Thirdness stands between Firstness and Secondness. Its facets include:
developing, growing, bringing about, representing. Thirdness bridges over the chasm
between First and Second, and draws their phenomena together. If we accept these
fundamental modes of consciousness as ontologies, they afford a psychological explanation for the three logical conceptions of quality, relation, and synthesis or mediation. This trichotomic, Peirce anticipated, would unite psychology, physiology, biology, physics and metaphysics and eventually serve as the phenomenal structure for a reorganization of all human knowledge around these universals. Although these three categories are not strictly pragmatic in the sense that we are unlikely to attain universal agreement on their character, they are necessarily prerequisite to understanding Peirce’s logic and his theory of signs. Peirce acknowledged that his categories bore some resemblance to Hegel’s thesis, antithesis, and synthesis, but he consistently rejected Hegel’s nominalism.

4.32 Semiotics and signs.

All thought is conducted in signs, Peirce held, and the study of signs is co-extensive with logic itself. He used the term *semeiotic* for the theoretical refinement of our common sense idea of a sign. Perhaps unsurprisingly for someone who spent 30 years with the geodetic survey living under canvas, working daily with pendula to triangulate and map new territory, Peirce conceived of a sign as an irreducible triadic relation. Semiosis involves a cooperation among (1) a Sign (sometimes known as a representamen), (2) an object, and (3) an interpretant. According to Pierce, when anything signifies or represents another thing, it does so in one of three ways: by its likeness or analogy with the thing signified (*icons*), by being physically connected therewith (*indices*), or by being associated therewith in the mind through habitual association (*symbols*). Through this third mode, people signify that a proposition represents a fact, but this mode cannot exist isolated from the other two: the tri-relative influence cannot be broken down into actions between pairs (EP2. 411). The universe which is the real subject of discourse can only be pointed to, never described. Therefore,
his conception of a sign as an indecomposable, non-severable triad departs from
European analyses of signs (from Aristotle and the Stoics to de Saussure) as dyadic
relations between signifier and signified. Figure 4.1 illustrates his triadic conception.
Peirce was explicit that the sign triad encompasses linguistics, graphics, music, and
mathematics (Tiercelin 2010; W5. 162-190).

![Sign Diagram]

Figure 4.1. Peirce’s triadic conception of the Sign.

For Peirce, every concept and every thought beyond immediate perception is a
sign. The object is not fully expressed in the Sign as a representation. Our purposes
always colour the conception of an object, through our consideration of effects. The
interpretant is the apprehended and total proper effect of the sign. The logical
interpretant does not directly correspond to any kind of object, but is essentially in a
future conditional tense, what Peirce calls a “would be.” Therefore, the logical
interpretant is in the nature of a belief-habit. The habit alone, though it may be a sign in
some other way, does not call for further interpretation but rather for action. “The
deliberately formed, self-analyzing habit . . . is the living definition, the veritable and
final logical interpretant” (CP5. 491). Therefore, the most perfect account of a concept
that words can convey will consist in a description of the habit which that concept is calculated to produce. A habit can be described by the kind of action to which it gives rise, with the specification of the conditions and of the motive. For Peirce, semiotics is not exclusively a mental activity; all conceptions will have anthropomorphic if not zoomorphic elements. He noted the involuntary response of a soldier to an officer’s order to ground his rifle, the flutists’ interpretation of a note of sheet music, the galvanic twitch of a decapitated frog’s leg, and the sunflower’s rotation to follow the sun across the sky (CP2. 274) as examples of habituated interpretants for a Sign\(^5\).

Another way of describing the relationship among the elements in his triad might be to say that we perceive the world as a diverse array of pure qualities or sensory impressions. Peirce would consider each of these qualitative entities of perception as a First, or Sign. Each such perception is conceptualized as “something,” a Second or object. Furthermore, each individual perceiver conceives the perceived Sign as an object according to his or her own personal “habit of mind” shaped by experience, a Third or interpretant. As he noted,

All our knowledge may be said to rest upon observed facts. It is true that there are psychological states which antecede our observing facts as such. Thus it is a fact that I see an inkstand before me; but before I can say that, I am obliged to have impressions of sense [Firsts] into which no idea of an inkstand, or of any separate object, or of an “I” or of seeing, enter at all; and it is true that my judging that I see an inkstand [Second] before me is the product of mental operations [Thirds] upon these impressions of sense. But it is only when the cognition has

---

\(^5\) Some might argue this is not different from Plato’s implication that true belief is more important than fact, because we act on what we believe, regardless of the “fact”. Yet Peirce had many quibbles with Plato, arguing that his philosophy was too poetic, neglected external causes, over-emphasized dichotomy, and wrongly convinced pure mathematicians that they could pursue ideas as abstract demonstrations without revealing the inner workings of the human mind as it confronted facts (EP2. 37-40). He equated Platonism with the *a priori* method for fixating belief. For example, he noted that Plato—like many other metaphysicians who attempt to found their propositions on “agreeable with reason” rather than “agrees with experience”—thought it reasonable “that the distances of the celestial spheres from one another should be proportional to the different lengths of strings which produce harmonious chords.” (EP1. 118-119)
become worked up into a proposition, or judgment of a fact, that I can exercise any direct control over the process. (CP6. 522)

In other words, the First is a perception (such as a statistical Sign), the Second or object is the conception derived from that perception (such as student achievement), and the Third or interpretant is the “habit of mind” of the interpreter who holds a conception (such as an average student performance). Within Peirce’s sign relationships, he distinguished between “what we are immediately conscious of” (the Sign) and “what we are mediately conscious of” (the object). Prima facie, we may be more inclined to consider the object (for example, student achievement) as the First, and the Sign as the “mind’s conception,” (for example, the statistical average) rather than the other way around. Indeed, Peirce’s insistence that the Sign rather than the object is the First in his triadic formulation, may appear incongruous and difficult to grasp. Yet it renders his semiotic theory of interpretation uniquely flexible in straddling both the realist and constructivist paradigms in sociology (Burrell & Morgan, 1994): a sign is a synthetic judgement and permits ampliative reasoning. This prevents his pragmatism and any analytic conceptions from being defined fully as realism (which becomes subject to positivist reduction) on one hand, or construed as an ungrounded, purely subjective idealism (and therefore largely inaccessible to empiricist verification and agreement) on the other.

Figure 4.2 illustrates how any given perception (of a Sign) may be conceived differently (as an object) by individuals (personal interpretants) who may differ not only in their education and experience, but also in other ways. Their social traditions, mathematical prowess, occupational position, circle of professional relationships, and other experiential characteristics may all contribute to distinctive “habits of mind” or

---

6 Some alternate descriptors that Peirce used late in his career for his brand of pragmatism were “critical common sensism” (EP2. 346-354), “scholastic realism” (EP2. 354), and “a form of conditional idealism” (EP2. 419; CP5. 494).
Thirdness, and thus to the ways they conceive what the average means. Peirce’s model of sign relationships therefore has many implications for organizational analyses, for school principals seeking consensus, for understanding conflict, and for engaging family and community in student achievement issues. Within the assumptions of this model, Figure 4.2 provides parameters for the management of meaning for an average.

An example demonstrates the way in which a single perceived Sign (at the centre of the diagram in Figure 4.2) might be conceived differently by three different individuals also represented in the diagram. If the Sign were written as 66 percent on a report card, the teacher who generated the mark might conceive it as two-thirds mastery of a curriculum guide’s outcomes in a flow conception of the average, but has computed it as a weighted mean. The school principal reviewing the mark might conceive the sign as indicating that the student ranks in the middle of student classroom performances, using a point-estimation conception of the average, but drawing on the median.

![Diagram](image)

Figure 4.2. Sign as perceived and conceived by three individuals with dissimilar backgrounds.

The student might initially feel elated that the 66 percent final mark as a quality or First, is a substantial improvement over the 55 percent attained at midterm, but hold a multiplicative conception for the average as an unweighted arithmetic mean. A fourth individual, the university registrar might consider the notation as signifying inadequate performance for university entrance purposes, holding a typical or representative notion of the mean. For the parent, the report card average might be conceived as a message about the student’s study behaviours in a turbulent household, holding a signal-in-midst-of-noise conception.

Moreover, the 17-year-old boy holding that report card mark, perceived as a Sign, may be interpreted as a friend, a student, a member of a school community, a family member, a potential university candidate, or as the product of the teacher’s, the school’s or the school division’s performance—all are different objects. In addition, another individual without basic numeracy skills, or from a different cultural background, may not be able to meaningfully conceive the notation of 66 percent at all, or identify any of the social attributes of that 17-year-old boy. In no case can the Sign be truly conceived as what it is in a definitive sense, since any such description would inherently limit the sign to the way in which it is metaphorically conceptualized as an object by one interpretant. Rather, the sign’s meaning inheres within the habits of mind of an individual in terms of its effects, which sits at the root of our problems in interpretation described in Chapter Two.
### 4.3.3 Sign types and classes.

Like many Victorian scientists, Peirce was concerned with taxonomizing within and among the emerging sciences. In his ever-trifurcating cosmos, Peirce identified three types of signs that can be recombined into nine classes according to their function.

Importantly, Peirce believed that circumstances must be considered in functional analyses of signs; context itself is a particular constellation of signs. It is not necessary to be accurate or precise in one’s analysis for the ordinary purposes of logic. These relationships are delineated in Figure 4.3 below.

<table>
<thead>
<tr>
<th>Sign (or Representamen)</th>
<th>Relationship of Sign to Object (Second)</th>
<th>Relationship of Sign to Interpretant (Third)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>First</strong></td>
<td><strong>(Qualisign)</strong></td>
<td><strong>(Icon)</strong></td>
</tr>
<tr>
<td>An embodied quality that is a Sign (i.e., whiteness, hardness, first, second, third)</td>
<td>A Sign that refers to the object it denotes by virtue of its characters [or sensory qualities of its own] . . . whether any such object exists or not (i.e., a piece of the cross, a map, onomatopoeia; an asterisk; an equality sign; the exponent 3).</td>
<td>A Sign of qualitative possibility, understood as representing such and such a kind of possible object. (i.e., morphemes in language, prefixes and suffixes, lines and boxes in a confidence interval, a pointing finger in a gesture, the natural number system.</td>
</tr>
<tr>
<td><strong>Second</strong></td>
<td><strong>(Sinsign)</strong></td>
<td><strong>(Index)</strong></td>
</tr>
<tr>
<td>An existing thing or event (i.e., a bridal gown, a fight, a trimester in a pregnancy)</td>
<td>A Sign that refers to the object that it denotes by virtue of its being “really affected” by that object (i.e., smoke and fire, red litmus and acidity, a symptom and its cause; ordinal numbers such as the third degree).</td>
<td>A Sign of actual existence (i.e., sentences, propositions, judgements, a diagnosis, an algorithm or equation)</td>
</tr>
<tr>
<td><strong>Third</strong></td>
<td><strong>(Legisign)</strong></td>
<td><strong>(Symbol)</strong></td>
</tr>
<tr>
<td>A social law or convention that is a Sign (i.e., marriage, a triangle)</td>
<td>A Sign that refers to the object that it denotes by virtue of a law, usually an association of general ideas, which operates to cause the symbol to be interpreted as referring to that object (i.e., a word, a sign language gesture).</td>
<td>A Sign of law, (i.e., the truth or reality of a given cultural group, a mathematical proof, Newton’s laws; ( E = mc^2 ), The Holy Trinity)</td>
</tr>
</tbody>
</table>

*Figure 4.3.* Taxonomy of Peirce’s Sign relationships.
We must underline that only one dimension of a perceived Sign is conceived (as an object) according to the “habit of mind” of an individual person (interpretant) at a given instant. Therefore, every sign has the potential to be simultaneously perceived and conceptualized in many different ways by different individuals. Anthropologist E. Valentine Daniel (1984) has offered a cogent simile for describing the multivalent nature of Peirce’s sign:

A . . . sign can be likened to a jewel . . . . In the experience of human interpretants, often only one facet will tilt toward the observer-interpreter, catch light, and throw it back; the other facets merely refract light. The reflecting facet, then, becomes the dominant mode. This does not, however, deny the existence of other modes of signification; modes that are sometimes expressed subjunctively and at other times remain potentialities. Different contexts help bring to light different aspects or modes of the sign. (pp. 39-40)

Overall, the most innovative element of Peirce's semiotic was his conception of the interpretant. Peirce used the term 'interpretant' “because it fulfills the office of an interpreter who says that a foreigner says the same thing which he himself says” (EP1. 5).

Every sign “lives” through its translation in a virtually infinite series of interpretants:

A notation on an individual student’s report card may thus be reconfigured as average student marks for a school or for a province, and reinterpreted by a university admissions officer, an employer, or a biographer decades later. Eventually, Peirce distinguished between three types of interpretant: the immediate, the dynamical, and the final. The final interpretant is the ideal interpretant toward which the semiosis tends under favourable conditions, or “the effect that would be produced on the mind by the sign after sufficient development of thought.”
The Final Interpretant does not consist in the way in which any mind does act but in the way every mind would act. That is, it consists in a truth which might be expressed in a conditional proposition of this type: 'If so and so were to happen to any mind this sign would determine that mind to such and such conduct’. By conduct, I mean action under an intention of self-control’ (CP8, 315).

The interpretant must therefore be understood, then, not as an act of interpretation (which approximates Peirce’s notion of conception), but as the sign's “proper significate effect” on the reader.

4.34 Sign transformation and inference.

Peirce held that Signs such as the average—and our interpretations of them—are continuously in motion and evolving. Through the process of semiosis, which Peirce described as the “action or influence” of signs upon one another, signs typically change.

Figure 4.4 illustrates how any given perception (as a Sign) may eventually be conceived and reconceived differently (as an object) by a single, continually changing individual (an interpretant) over a person’s lifetime. Learning is the transformation of these signs, since the Sign is the medium for communicating a form.

Figure 4.4. Evolving interpretant over time, resulting in changes to the conception of the Sign and its object.
Upon an individual’s first conscious perception of a given Sign, the sign may be regarded as new data, with potential meaning to yield. However, as she or he encounters or reflects upon the same Sign over a period of time, placing it in a variety of other statistical, educational or social contexts, the sign may become more meaningful—more well understood and more well-integrated into the individual’s habitual patterns of thought as beliefs. A favourite story, a religious ritual, a statistic, a friend, an enemy, or a spouse will gradually yield less and less information to a person, but become a more meaningful part of life over time as his or her habits of thought and action grow and change. The way the individual conceives the interpreted Sign with its object is thus likely to evolve over time. We must also underline that as the individual him- or her-self (considered as a Sign) changes throughout this process of semiosis, the manner in which he or she is likely conceived (as an object) by another interpretant (such as a friend or spouse) is also likely to change. What is deemed as a statistically average school or university on one day may be interpreted quite differently on another.

Peirce’s triadic classification of signs (icon, index, and symbol) can be combined with his triadic model of inference-making as forms of reasoning (EP2. 267-288). Of course, as noted above, inferencing includes abduction (begins with experience and ends with adoption of another sign; and generation or appropriation of signs to explain novelty); deduction (begins with existing signs according to prescribed syntactical rules, with no new ideas admitted to the sign system); and induction (begins with a well-articulated sign system, and applies it with rules to specific experience so it becomes more understandable in terms of the existing sign system). In semiosis, the processes of abduction, deduction and induction are non-linear, unselfconscious, and enmeshed with our deliberate inventive capacity.
To describe these processes, Peirce further reassembled his taxonomy of sign types into ten *classes* of inferencing (CP2. 233-72; EP2. 294-299). Mathematics educators (Kehle & Cunningham, 2000) have updated Peirce’s terminology into contemporary terms for ease of understanding and application:

- **Hunches** (rhematic iconic qualisign) involve abductive reasoning to determine whether the initial observation has enough properties that might lead to possible evidence; an *omen* is a sign whose resolution occurs in future acts of inquiry, whereas a hunch is an implicit inference of an omen.

- **Symptoms** (rhematic iconic sinsign) involve abductive reasoning toward possible resemblances, to ascertain whether an actual observation has enough properties to be case relevant; a symptom is a sign whose action is ongoing in the present. We infer from the symptom the presence of a more general phenomenon.

- **Metaphors or analogies** (rhematic iconic legisign) are modes of abduction involving the manipulation of resemblance to create or discover a possible rule of order.

- **Clues** (rhematic indexical sinsign) exemplify abductive reasoning backwards in order to determine whether some observations are clues of a more general phenomenon; unlike a symptom, the clue is a sign of some past state of affairs.

- **Diagnoses or heuristic thinking** with scenarios (rhematic indexical legisign) is a form of abductive reasoning about a possible rule based on available evidence, proposing plausible hypotheses or scenarios from the body of clues; less emphasis is placed on whether the clues are relevant and more on how relevant.

- **Explanations** (rhematic symbolic legisign) involves abductively asserting a possible rule, or reasoning in order to form a general plausible explanation. If inferences have led us to a hypothesis that can be formally stated as a general rule, then it is ready to be linked with other rules via deduction and tested via induction.

- **Identification** (dicent indexical sinsign) is a form of inductive reasoning that involves testing for actual evidence of a particular thing, whether our observation is an instance of $x$. This is scientifically known as construct validation and associated with thought experiments.

- **Prediction** (dicent indexical legisign), in contrast, is an example of inductive reasoning from actual evidence of a possible rule; when constructs are linked in some causal or covariant relationship, it is more formally known in scientific terms as hypothesis testing, or in administrative terms as planning.

- **Model-building** (dicent symbolic legisign) involves inductive reasoning where tests lead to a probable conclusion based upon a rule or set of rules to build models. When rules form a coherent whole and create a structure from which
actual experience can be tested, then habits, models or world views emerge; in scientific parlance, this sort of inquiry is called convergent validity.

- **formal reasoning** (argument symbolic legisign) involves deductive reasoning where a necessary conclusion is reached based on formal rules. A problem-solver might link hypotheses for further inductive and abductive reasoning.

Each of the six modes of abduction deal with potential or possibility. The three modes of induction deal with actuality or probability, and the deductive mode focuses on necessity or rules and regulations. In the procedural rush of semiosis, people may employ all three modes in dynamic interplay with each other, engaging qualities, actual things, or generalities depending on the type of sign. Metacognitive thinking involves awareness or monitoring of one’s own reasoning patterns. Viewed semiotically, metacognition occurs when a past instance of semiosis becomes the signs and/or objects of a present instance of semiosis. All such inferencing is subtle, fluid, recursive, and reflexive.

Thus, Peirce’s theoretical framework allows us to discern and describe school principals’ reasoning patterns when interpreting statistics in a relatively precise manner: we are not just performing philosophical calisthenics, nor simply cataloguing with a new lexicon. However, in removing boundaries from the notion of “community,” his semiotics creates ambiguities in defining what may be “generalizable” in multiple senses of the term. Peirce was aware of this issue. He insisted on plausibility as the central criterion when reviewing knowledge claims, rather than “transferability,” as do contemporary knowledge mobilization theorists, or “generalizability,” as do cognitive scientists.

### 4.35 Convergence and norms.

In a pragmatic outlook, normative processes are inherent in all phenomena and inescapable. In Peirce’s definition, normal is “not the average (or any other kind of
mean) of what actually occurs, but of what would, in the long run, occur under certain circumstances” (CP6. 327-8). Not surprisingly, he had verified this assertion well before other researchers in the nascent field of statistics through an extended experiment involving perception of sound (Peirce, 1873; Wilson & Hilferty, 1929). Unlike many contemporary researchers, therefore, who presume they can step outside to superimpose a norm on their experiments post hoc with a Gaussian curve, Peirce believed that the sciences themselves, logic, and human observation in particular, are intrinsically normative. (We perceive that which we are conditioned to perceive). For Peirce, reasoning occurs primarily within a social setting, and Reality is ultimately shaped in the long run, within an indefinite community. All knowledge has a social character, and cannot depend on the inferences of single individuals, for individuals die:

Now the number of risks, the number of probable inferences, which a man draws in his whole life, is a finite one, and he cannot be absolutely certain that the mean result will accord with the probabilities at all. Reasoning inexorably requires that our interests shall not be limited. They must not stop at our own fate, but must embrace the whole community . . . . He who would not sacrifice his own soul to save the whole world, is, as it seems to me, illogical in all his inferences. Logic is rooted in the social principle. (CP2. 645-60)

Therefore, for Peirce, the idea of a transcendent community supersedes the individual purposes of a single reader or author of statistics. It is not the structure of an individual human mind that produces the interpretation, but rather the reality that semiotic acts build up within the collective mind.

From the moment in which any community is persuaded to agree with a given interpretation, there is, if not an objective, at least an inter-subjective meaning that acquires privilege over other possible interpretations articulated without the agreement of the community. Peirce was explicit that the community of researchers is independent of what we as an indefinite group may think. The purpose of all scientific investigation is to
push the collective opinions of everyone in the world closer and closer to agreement, and thus closer to the limit represented by reality, toward a common core of ideas. The true conclusion of semiosis is Reality. There is an ideal perfection of knowledge, in the long run, toward which normative judgements and our conceptions of the average contribute. The thought or opinion that defines reality must therefore belong to a community of knowers, and this community will be structured and disciplined in accordance with supra-individual principles. The Real is the idea around which the community ultimately converges. The opinion that is fated to be ultimately agreed upon will be shared by all who investigate and is what we mean by Truth; the object represented in this opinion is the Real.

The [R]eal, then, is that result which sooner or later, information and reasoning would finally result in, and which is therefore independent of the vagaries of me and you. Thus, the very origin of the conception of reality shows that this conception essentially involves the notion of a COMMUNITY [capitals in original], without definite limits and capable of an indefinite increase in knowledge. And so those two series of cognitions—the real and the unreal—

---

7 An analyst-wag in a government evaluation branch where I once worked described it as the “Department of Non-Fiction.” He once jestingly referred to me as “Fact-tote-him” and often asked which book I wanted to borrow that day on two-week loan.

8 Among the many different approaches to “Truth” that philosophers adopt—a copy theory, a coherentist perspective, a correspondence theory, the convergence view—Peirce’s most closely resembles the convergent: the predestinate opinion which would ultimately prevail if investigation were carried sufficiently far (C2.457; CP1.139). However, he underlined that “the further investigation is pushed, the greater will be multiplication of unsettled questions” (CP2.457). He disagreed with James’ view of truth as a satisfactory state of affairs (CP2.379) because of its nominalism, preferring to describe truth as the conformity of a representation to it’s object through an interpretant (CP2.304, 380). Peirce acknowledged that the concept of truth differs in mathematics, in ethics, and in the sciences (CP2.87). However, although mathematics supervenes ethics, aesthetics, and the sciences in his philosophy, he argued that mathematics cannot represent the truth of feeling in its iconography. Peirce often enclosed the word in quotation marks, in both the upper and lower case, insisting that our problems would be greatly simplified if we defined Truth as “a state of belief that was unassailable by doubt” (CP2.336). Of course, Peirce’s definitions were more sophisticated in the realm of mathematics, where he is recognized as one of the first to construct truth tables. His “On the Algebra of Logic: A Contribution to the Philosophy of Notation” in the American Journal of Mathematics of 1885 includes an example of an indirect truth table for the conditional.
consist of those which at a time sufficiently future, the community will always continue to reaffirm; and those which, under the same conditions, will ever after be denied. (CP5. 311)

In other words, Peirce believed that with sufficient computation and consideration of the meaning for various averages of student achievement, the community will eventually agree on what the Real middle is in student prowess.

4.36 Causation and consequences.

Notions of causation, effects and consequences are central to understanding models of instructional leadership, the interpretation of statistics, and the management of meaning. “Agency” refers to the capacity of individuals to act independently and to make their own free choices, whereas “structure” refers to those factors (such as social class, religion, gender, ethnicity, subculture, etc.) that seem to limit or influence the opportunities that individuals have. Human agency is characterized by intentionality and forethought, self-regulation and self-reflectiveness about one's capabilities. Personal agency operates within a broad network of socio-structural influences. In these agentic transactions, people are both producers and products of social systems. Current social cognitive theory distinguishes three modes of agency: direct personal agency; proxy agency that relies on others to act on one's behalf to secure desired outcomes; and collective agency exercised through socially coordinative and interdependent effort (Bandura, 2001). Our causal language—words like “influences,” “determines,” “shapes,” “drives,” “breaks,” “impacts”—expresses the agency of the school principal as reader of statistics, and indicates how or if the reader will proceed to bring about changes.

Unlike many current scientists who are concerned about the nature of the relation between cause and effect, Peirce focused on the relata (the categories of events) involved. As in many aspects of his thinking, Peirce starts not with cause-and-effect in the physical sciences, but with antecedent-and-consequent in logic. There is a difference
between saying a student’s average achievement is caused by poor study habits, and saying that the student has average student achievement as consequent to not submitting assignment A and performing poorly on examination B. The former is a claim for a causal relationship, whereas the latter is an inference with evidence. Peirce did not consider causality as either a primordial element of nature or a fundamental category of thought, but rather as a type of retroductively-inferred facts (CP6. 66). Our conceptions of causality are contradictory, he believed. All signs are associations of ideas that consist of perceptual judgements, where one judgement occasions another judgement of which it is a Sign. This is nothing more or less than inference (CP5. 264-317).

Peirce considered causation as tripartite, having final, efficient and chance components. His conception of Final Cause cannot be adequately understood outside the scientific and theological debates underway at Harvard and elsewhere in the latter 1870s—debates fuelled by Darwinian biology, the higher biblical criticism, and the Unitarianism of Peirce’s family. Secularization, I believe, is the source of the scholarly confusion and ambiguities found in the (non-)capitalization of Final Cause in Peirce’s writings. Unlike theologians, Peirce’s central contention was that final causation is explicitly and intentionally anthropic, if not anthropomorphic. At issue is whether school principals, when reading statistics, believe “average student achievement” can be effected by their leadership behaviour in the immediate term, whether they believe that their leadership actions are part of a larger pattern of causation, whether they position themselves as largely constrained by the capabilities of students and staff in their school, or whether they react to achievement as happenstance, as a form of chance.

Anticipating the Gestaltists, Peirce defined efficient causation as where the parts compose the whole, and final causation as where the whole calls out the parts (CP1. 220). Final Causes are not future events, but general (physical) possibilities that may be
realized in the future, an ideal end state toward which a process tends. Unlike final causation, efficient causation is not directed toward an end in any way. Peirce offered as an example someone who intends on shooting a bird. To wing the bird, he aims slightly ahead of the bird, taking into account the distance the bird will fly before the bullet reaches it. This activity is end-directed, and thus belongs to final causation. But as soon as the bullet leaves the rifle, there is only the unintelligent, efficient causation that is in no way concerned with the results of its activity; the bullet will not follow the bird swooping in another direction. Efficient causation has no regard for results, but simply follows orders blindly (CP 1.212). A school principal who acts on a student report card mark because it is board policy is an efficient causal agent; one who purposively acts because s/he believes their intervention will make a difference is a final causal agent; one who encounters a student in the hallway and happens to mention something about a mark is operating by chance.

Peirce also expressed these interrelationships in a courthouse analogy:

The court cannot be imagined without a sheriff. Final causality cannot be imagined without efficient causality; but no whit the less on that account are their modes of action polar contraries. The sheriff would still have his fist, even if there were no court; but an efficient cause, detached from a final cause in the form of a law would not even possess efficiency” (CP 1.213). . . . Final causation without efficient causation is helpless. . . . Efficient causation without final causation, however, is worse than helpless, by far; it is mere chaos; and chaos is not even so much as chaos, without final causation it is blank nothing. (CP 1.220)

In summary, according to Peirce, final causes are synonymous with beliefs: they habitually direct processes toward an end state. Like human habits, habits of nature (laws of nature) are final causes because they display tendencies toward an end state. Final causes stand to laws of nature as genus to species. Moreover, habits are not static entities, for they may evolve over time. Peirce called the possible evolution of final causes “developmental teleology,” similar to cognitive anthropologists’ notion of
“emergent goals.” For any study that looks at the causal agency of the reader of a Sign such as an average, a fundamental issue is whether they partake of efficient or final causation.

Can the sign of an average “cause” a behaviour? A debate has emerged among semioticians about whether causality is intrinsic or extrinsic to Peirce’s semiotic theory. For some scholars (Short, 1981, 2007), the term “determined” means “caused,” whereas for others (Hulswit, 2002), causal relations are not inherent in Peirce’s triadic model of a sign although he often used the term “caused” in his notion of Secondness. My own view is that we must remain skeptical about causal relations as intrinsic to a sign triad because (a) Peirce always insisted that the triadic relation was irreducible, whereas imputing causality to any one of its dimensions severs that relationship; (b) he believed that semiosis was intrinsic to nature, and he was definitive in stating that causality was not, and (c) Peirce tended to see causal relations as enacted over a long duration in his continuity and synechism, rather than within the instantaneous perceptions of sign-readers. Perhaps the best way of clarifying Peirce’s intentions is to return to the distinction between a Sign (as notation on paper), in which consequential implications are inherent in the mind during conception, and a sign (as notation, object and reader’s interpretant together), where effects may causally follow when the reader acts upon it.

This issue is important for studies of perception, conception and statistical understanding, because it interposes a causal relationship within any controlled study between the “gates” of perception and action with respect to averages. To my way of thinking, at issue are fine distinctions between “cause” and “because,” the former suggesting effects and the latter suggesting logical consequents. Cause suggests inductive inferencing, whereas “because” may embrace abductive reasoning. Peirce remained elusive about the meaning of “determined” in his theory of signs. Writing late
in life to his benefactor, William James, Peirce summarized much of his semiotic by declaring that

a Sign is anything of either [sic] of the three Universes being bestimmt by something other than itself, called its Object, in its turn is bestimmt the mind of an interpreter, to a notion which I call the Interpretant; and does this in a manner that the Interpretant is thereby and therein, determined mediately by the Object. (EP2. 497)

Peirce acknowledged that the term bestimmt could be construed as meaning “caused.” However, he pointed out that a sign could not really be causal in nature because hygrometric signs by indicating future weather as an unreal object could not really cause a person to retreat indoors. Perhaps, the most we can say is that in Peirce’s theory of interpretation, semiotics does not presuppose causal relations, but causation does presuppose semiotics as meaning-making with signs.

4.4 Recapitulation

Peirce’s theory of interpretation provides an all-encompassing, yet incisive and internally coherent set of ideas for a complementary pair of investigations into school principals’ conceptions of and practices involving average student achievement. Based on a wide but unavoidably incomplete review of Peirce’s writings, we may conclude that:

1. The three phenomenal categories of Firstness, Secondness, and Thirdness are central premises in Peirce’s theory of interpretation. Our conceptions of an average may thus be studied as a phenomenological form which embraces: (1) apprehended qualities; (2) a set of reactions to brute facts, and; (3) a mediative resolution.

2. An average is a form of sign and thus implicated in any semiotic process. Every thought or sign that registers in consciousness involves an indivisible triadic relationship between the Sign (a First, which is a perception), an object (a Second, which is a conception derived from the perception), and an interpretant (a Third, which is the effect or consequent in the mind of the perceiver which determines how the perception will be formed as conception according to a habit of mind).

3. Different individuals may conceive of the same Sign differently in relation to an object, because of differences in “habits of mind.”
4. Only one dimension of a perceived sign is conceptualized (as an object) by an individual according to their “habit of mind,” although every sign has the potential of being perceived and conceived in many different forms (i.e., icon, index, symbol) and in different ways over time.

5. Because Peirce was interested in concepts of chance and coined the phrase normative distribution, we must attend to allied conceptions of variation, distribution and proportion when considering a school administrator’s interpretation of an average.

6. A statistic or average is a Sign, or a constellation of Signs in a formula, to which divergent interpretations may be brought, and for which varying objects can be assigned.

7. Although Peirce’s terminology for various sign forms is dated, contemporary scholars in mathematics and semiotics have successfully updated his vocabulary, which enables us to distinguish between various forms of inference when interpreting an average.

8. Peirce was ambiguous about whether any sign such as an average causes or induces the interpreter to adopt a particular interpretation.

The foregoing propositions, taken in combination with Peirce’s pragmatic ideas described earlier, furnish both the philosophic and methodological foundation for this dissertation’s design.
To develop its meaning, we have, therefore, simply to determine what habits it produces, for what a thing means is simply what habits it involves. Now, the identity of a habit depends on how it might lead us to act, not merely under such circumstances as are likely to arise, but under such as might possibly occur, no matter how improbable they may be. What the habit is depends on when and how it causes us to act. As for the when, every stimulus to action is derived from perception; as for the how, every purpose of action is to produce some sensible result.

Charles Sanders Peirce

Chapter 5: Methodology and Methods

The sequential, exploratory, mixed methods plan for this inquiry into school administrators’ conceptions of the average is outlined in this chapter. Two methods were used. The qualitative method was based on Peirce’s precepts about phenomenal inquiry, and the quantitative method was founded in his guidelines for drawing statistical inferences within randomized experiments. Both methods draw on the pragmatic precepts of Peirce’s philosophy, which in this study is considered as a methodology.

5.1 Presuppositions

North American pragmatism straddles the constructivist and empirical paradigms. Some have argued that constructivist and positivist ontologies are irreconcilable. According to Lincoln and Guba (2000), positivism's naive realism holds that educational affairs are both real and apprehensible whereas in constructivism, meaning is generated by individuals and within groups. This analysis implies that corresponding qualitative and quantitative research methodologies, respectively, are incompatible. However, researchers (Cupchik, 2001; Paul, 2005; Rennie, 2000) now recognize that classical pragmatism may embrace both constructivism and positivism in a form of constructive realism. The first step is to acknowledge a social world or worlds that are reflected in the natural rhythms of daily life that exist prior to and independent of either positivist or
constructivist analysis; hence realism. Phenomena must be understood as thinking processes that cut across the physical, social, and personal (self) worlds (Peirce & Ketner, 1992, pp. 146-149). Averaging is one such process, and hence conception as the act of interpreting a Sign is partially accessible via either phenomenological or experimental methods. Both qualitative and quantitative researchers may examine the same phenomenon, offering complex descriptive accounts or precise analyses of functional relations, respectively. I presuppose that both approaches face the problem of constructing data and are therefore subject to potential bias. While description has traditionally been viewed as preceding explanation, (i.e., natural history precedes hypothesis testing), the two approaches are viewed in this dissertation as complementary and may be conducted in parallel or serial order. Qualitative methods offer an in-depth account of underlying perspectives and concepts to identify categories and potential functional relationships. Quantitative methods yield empirical findings related to changes in processes and can suggest areas that might further be elaborated through detailed examination and description. Such pragmatism as espoused by Peirce underlies mixed-methods designs (Tashakkori & Teddlie, 2003, 2010).

Such an approach is supported extensively in pragmatic philosophy. Peirce distinguished between conditional possibility and conditional probability when appraising habits, as routinized behaviours and thoughts that intersect (EP1. 131). “Possibility” as a First logically precedes “probability” as a Second in his pragmatic philosophy. As Peirce noted, “… it seems undeniable that there really are such possibilities, and that, though they are not existences, they are not nothing. . . . . But whether this be admitted or not, it is undeniable that such elements are in the objects as we commonly conceive them; and that is all that concerns phenomenology” (EP2. 269). In other words, what “may be”
comes before what “would likely” happen, and the preferred sequence for the phases of inquiry becomes evident (EP1. 248-250). Therefore, the researcher must draw on conjecture or retroductive thinking for guiding qualitative inquiry, while preferring inductive and deductive thinking to test hypotheses in quantitative inquiry. Peirce was explicit that his methods apply equally as well to the interpretation and analysis of historical documents and testimonies (EP2. 75-114) through balancing likelihoods. It follows, then, that two different methods may be used to explore school administrators’ conceptions of the average.

5.2 Permissions

Approvals from the University of Regina’s and the University of Saskatchewan’s Research Ethics Boards were secured in November 2012 to March 2013 for both the qualitative and quantitative components of this inquiry, through an inter-university protocol (see Appendix A). In all cases, informed consent was obtained from respondents in the study’s qualitative phase, and from participants in the quantitative phase. In the qualitative phase, a consent form was signed by interview participants before undertaking the research (see Appendix B). In the quantitative phase, informed consent was deemed to have been obtained with submission of the completed survey. Except for one occasion, when double-booking necessitated that a third party researcher introduce and collect paper surveys, I was present on every occasion when data were collected from those participating in this project. For the quantitative phase, I secured written permission in spring 2012 from Dr. Seymour Epstein at the University of Massachusetts (Amherst) to use the Rational-Experiential Inventory (see Appendix C). The overall, mixed methods research plan is outlined in Figure 5.1 below.
Figure 5.1. Mixed-method research plan for sequential, exploratory inquiry into school administrators’ conception of the average: procedure and products.
5.3 Qualitative Method for Phase 1

5.31 Assumptions.

A central assumption throughout this dissertation is that to adequately evaluate the applicability of Peirce’s theory of interpretation and significance, the researcher must in turn adopt his methods for conducting inquiry. Therefore, I adopted the precepts of Peirce’s phenomenology for the first phase of this study. Peirce’s phenomenology is a research method that was foundational to the normative sciences as he conceived them, and was integral to his ontological stance. Although Peirce’s phenomenology was developed more than a century ago, it is still practical today. In a 1903 letter to William James, Peirce defined phenomenology as

the analysis of what kind of constituents there are in our thoughts and lives,(sic)(whether these be valid or invalid being quite aside from the question). It is a branch of philosophy I am most deeply interested in and which I have worked upon almost as much as I have upon logic. (CP8. 295)

Phenomenology, or phaneroscopy as Peirce sometimes called it, is a descriptive rather than an explanatory science that studies consciousness or subjective experience; he informed William James that phenomenology was entirely distinct from psychology (Taylor & Wozniak, 1996). A phenomenon, for Peirce, entails “all that is present to the mind in any sense or in any way whatsoever, regardless of whether it be fact or figment. In examining the phenomenon, we endeavour to sort out its elements according to the complexity of their structure, thus reaching three categories” (CP8. 213), which in turn will reflect First, Second and Third as his three architectonic categories, and his semiotic classification of signs. As he noted in correspondence: “A sign in itself may be an indefinite possibility, when I term as a Qualisign, or it may be an existent thing or event, when I term it a Sinsign . . . or it may be a general type, when I call it a Legisign.” The three categories are supposed to be the indecomposable elements that attentive perception
can make out in the phenomenon (CP8. 264-5). For this study, these elements are the essences of the average as interpreted by school administrators over three points in time.

One might misconstrue his phenomenology as exclusively mental, and as disregarding somatic or behavioural considerations. For someone as fiercely anti-Cartesian as was Peirce, such a stance is untenable. As previously outlined in Chapter 3, Peirce argued that mind encompasses the body and other individuals’ minds. Therefore, Peirce’s phenomenology extends beyond considerations of the individual psyche. Moreover, he was emphatic that phenomenology can take for scope anything, regardless of whether it corresponds to any real thing or not. His phenomenological writings do not prescribe particular times for study, such as before or after reading an achievement report or particular time in the school year. Nor does his phenomenology delimit particular individuals whose minds may be explored, leaving those questions to the researcher. He did not doubt that those universal features of the phenomenon that were delineated would also be present in all other minds and at all times. And he pointed out that, “I do not limit the reference to an instantaneous state of consciousness; for the clause ‘in any way whatever’ takes in memory and all habitual cognition” (EP2. 362). Insofar as Peirce elaborated this science of observation, phenomenology is concerned with the formal elements of the phenomenon. He knew that there was another series of elements imperfectly represented by Hegel's idealistic categories. However, he could never provide a satisfactory account of them (CP1. 284).

Phenomenology, in Peirce’s architectonic, involves the plausible depiction of categories, not in a statistical sense and not a purely experiential sense, but rather in a semiotic sense. Peirce indicated that, procedurally, the researcher must initially (a) identify several very broad categories, as an act of classification according to form. The researcher then (b) describes the features of each class. Thereafter, the researcher
will (c) demonstrate that although these features are inextricably mixed; no one set of features can be disassociated from the other. His phenomenological method culminates in (d) drawing up a very short list that comprises the broadest categories of a phenomenon. Finally, the phenomenologist proceeds to (e) the laborious and difficult task of enumerating and describing the primary subdivisions of those categories (CP1. 286-287; Rosensohn, 1974).

At the risk of oversimplifying differences between Peirce’s and Continental European brands of phenomenology, we may characterize Peirce’s phaneroscopy as (a) aiming at the highest possible level of generality but not necessarily generalizability in its descriptions, (b) emphasizing the phenomenon rather than experience, and (c) situating the researcher in an objective rather than subjective stance.

For Peirce, the first point about generality implies that to specialize in human ways of feeling, perceiving or thinking would be too restrictive. The realities discovered through mathematical reasoning are universal and not merely human (although we express them in a human format, as it were). In the process of articulating its results (which it must do in order to count as a science), phenomenology may draw on mathematical ideas as the only means of generalizing from honest observation (EP2. 145-159). Pursuing essences does not necessarily predefine them as unitary. As Peirce explained,

A reader may very intelligently ask, How is it possible for an indecomposable element to have any differences of structure? Of internal logical structure it would be clearly impossible. But of external structure, that is to say, structure of its possible compounds, limited differences of structure are possible; witness the chemical elements, of which the “groups,” or vertical columns of Mendeléeff’s table, are universally and justly recognized as ever so much more important than the “series,” or horizontal ranks in the same table. Those columns are characterized by their several valencies . . . (CP1. 289).
The second point about focusing on the phenomenon rather than experience underscores that although the researcher may want to describe the conditions or environment in which a person is operating, the point of attention is average student achievement, not the socioeconomic or organizational context (Rosenthal & Borgeois, 1980). Peirce was clear that the researcher must assiduously refrain from all speculation about the relations between the phenomenological categories and physical facts, cerebral or otherwise. The researcher in a Peircean analysis must confine himself to honest, single-minded observation of a phenomenon’s appearance. The phenomenologist does not undertake, and carefully avoids hypothetical explanations of any sort. The inquirer simply scrutinizes the direct appearances of the phenomenon. Minute accuracy is to be combined with the broadest possible generality (CP1. 284-287).

The third point about an objective stance underlines that the researcher’s greatest effort is not to be influenced by any tradition, any authority, any reasons for supposing that such and such ought to be the facts, or any fancies of any kind (CP1. 287). Because Peirce’s semiotic is co-extensive with both his pragmatism as a methodology and an epistemology, a phenomenological approach to the interpretation of an average must involve the investigation of perceived consequences and their antecedents, not making extrapolations from the present into the future as forms of predicted effects. The reader of any phenomenological record must repeat the author's observations for himself, and decide from his own observations whether the author's account of the appearances is plausible or not (CP1. 286-287).

Peirce’s phenomenology thus proceeds from philosophic bases different than some other brands of phenomenology (De Tienne, 1993). It does not seek to reduce any phenomenon to a single essence but forecasts three. Peirce’s recognition of three modalities of being stands in contrast to a depiction of the inquirer as either being or
becoming: Peirce adds co-being and co-interpretation (CP2. 378-379). Similar to many other methods of inquiry, phenomenology requires a study of the complexity in structure through the analysis of categories, as expressive of people’s purposes. For Peirce, the researcher cannot bracket out his preconceptions, because of the impossibility of disassociating oneself from one’s past and the inability to sever mind; epoché (the theoretical moment where all judgements about the existence of the external world, and consequently all action in the world, is suspended), is thus not possible. Rather, our efforts are dedicated elsewhere: to erasing one’s own understanding of other research traditions or a priori conceptions of the meaning for a phenomenon (CP1. 284).

5.32 Research question and abductions.

Accordingly, the qualitative phase of this study was guided by the research question, “What are school administrators’ conceptions for the term ‘average’ student achievement?” “What” at the beginning of the question suggests that I sought a clear and concise description of the forms and substantive features of an average, through depictions of the interpretive process in reading averages, and from testimony about the functional uses of the average. In this phase, I adopt the Markovian principle of the “drunk man’s walk” in fuzzy logic. We can ascertain that our actions tend in an approximate direction, but we can never predict with certainty what the drunk man’s next step will be at any given point in time while he is getting home. The qualitative phase therefore depicts diverse conceptions of the average at three points in time during Winter 2012-2013, and in relation to three data displays, but does not presume such meanings preordain principals’ behaviour.

Thus, both the interviewer and interviewee were free to articulate a variety of temporal circumstances as conditional statements. Conditional possibilities may be either retrospective (I would have done this, in that circumstance) or prospective (I would do this, in this circumstance). Interviews ranged freely across time, so as to illuminate
alternate conditions as circumstances or contexts within which the average may be manifest. I concentrated on the “when” and “how” of averaging: when it would be computed, when the principal would act upon it, when it would have such-and-such an effect; how it would be computed, how principals would approach an issue raised by the average, how the average would impact on a student or parent or teacher. “Average” was construed broadly: a conception may be expressed verbally or in writing as a word or phrase, in mathematics as a number or letter, visually through colour and shape, through gesture as motion of the hands and fingers (Calvo & Gomila, 2008; Killeen & Glenberg, 2010; Lakoff & Johnson, 1999), or graphically as a line or geometric diagram. In other words, I allow for an extended mind in cognition (Rupert, 2009). For the qualitative phase of this study, the term “average” included all potential signs or graphic depictions, as proffered by the respondent, thereby permitting the respondent to convey any oral, written, mathematical, visual, or gestural conception of the average as phenomenon. Many potential conduits were thus provided for conveying meaning. For this phase of the study, a conception is defined as a belief or habit that provides a reductive idea about reality. A student is defined as a person who is or has enrolled in the school for which the administrator has legislated responsibilities.

5.33 Respondents and sampling.

Ten school principals were purposively selected for three one-on-one interviews each. Interviews were conducted between December 2012 and April 2013. Principals were chosen within the guidelines provided for a purposeful, intensity sample (Patton, 2002; Teddlie & Yu, 2007): respondents were anticipated to have sophisticated or extensive experience with the phenomenon of interest, without being aberrant or extreme in approach. As Patton (2002) indicates, intensity sampling requires that the researcher undertake some prior preparation and exploratory work to identify appropriate respondents. It is a representative sample in the sense that these 10 school principals are
forecast to provide information in an especially clear fashion about the phenomenon, without distorting it. All principals were working in circumstances that are typical of Saskatchewan’s educational scene.

The 10 school principals who participated in this study were selected in a systematic and deliberate way. Seven school divisions were identified in different areas of the province, each with different situations—rural, urban and on First Nations reserve. A northern location was not sampled, but a school principal with experience in more than one northern Saskatchewan school was selected to participate. I consulted with division office superintendents, consultants, and Directors of Education to identify school principals who met the following selection criteria: (a) mixture of experience levels, (b) male/female balance, (c) urban/rural parity, (d) public and Catholic school system balance, (e) mixture of elementary, middle and high school circumstances, (f) employment at an on-reserve First Nations school, and, (g) experience in a remote northern school setting. Two lists of 10 candidates who met these criteria were created. Then, I contacted serially each candidate on the first list by phone to obtain their verbal and eventually written consent to participate, thus preserving their overall confidentiality. All 10 candidates in the first list agreed to participate and the second list was not used. Pseudonyms are used for participant, school, division, and location names throughout the dissertation to preserve anonymity.

My respondents were:

- Craig Tremaine is principal of a rural K-12 school in central Saskatchewan. He taught in northern Saskatchewan for three years and has been a principal in several schools. Craig characterizes himself as someone who is “trying to be more and more of a systems thinker, that’s what I’ve been criticized as not being, but I’m still pretty concerned about, ‘My fish, my pond.’ I make sure that we can do the best we can, indicating that when you have a responsibility for a certain set of fish, you watch your fish and make sure they’re eating right.”
Kirby Klein is principal of Centennial School, a small K-12 facility in rural Saskatchewan. Originally from Saskatoon, a graduate of Yale University and a former teacher in Vietnam, Kirby has a background in mathematics and physics. He does not believe that educators can quantify average student behaviour. “You know, you describe the general behaviour of a person, of another person, put a number to it, I don’t think you can.”

Louise Simpson has been principal of Jean Sauvé Comprehensive School for two years, which has approximately 540 Grade 9 to 12 students and is located in a small city in southern Saskatchewan. Lois was born and raised in Moosomin, and taught elementary school in several locations before becoming a resource room teacher and elementary school principal for seven years. “When the word average comes to my mind,” Louise asserts, “no right or wrong answer, just average; well I go right to people because I’m not a math girl.”

Moira Gullickson was principal of a Grade 1-to-6 elementary school in central Saskatchewan. On maternity leave, she is interested in data-driven decision-making, and has been an administrator for six years. Moira is also a PhD student at the University of Saskatchewan and particularly interested in applications of information technology, for which she has received national recognition as educator.

Phillip Brunswick is a principal in the same school division as Kirby, but in charge of a K-7 elementary school. He has 15 years of experience teaching in a neighbouring small town school. He has seven years of administrative experience. Phillip sees averages has having different strands:” I mean there’s an academic average, but when you talk about an average student, ’OK, this student is a 75 average, and everybody in his class is a 75 average,’ might not necessarily give you the full picture of that student.”

Pammy Reid is a veteran principal of East Lake Public Elementary School, a Kindergarten to Grade 8 school with 291 students. Most students come from rural areas around the small city in which it is located. Pammy believes strongly in criterion-referenced assessment, and resists the imposition of averages in school affairs.

Peter Getz is a 36-year educator and principal of a Grade 7-12 school in a rural Saskatchewan town of approximately 2000 people. He has worked 24 years in a smaller town as teacher, including 12 years as principal, and another 12 years as teaching principal in the current high school setting. Peter believes “there is real pressure that everybody succeeds. I don’t think it’s acceptable anymore from anybody’s perspective that we have kids failing school, so how do you do that?” And as he acknowledges, “I’m at the end of the rainbow responsible for the achievement of the kids in the school. The buck stops with me.”

Rick Gabriel is principal of a small Grade 1-12 First Nation-controlled school with 182 students on reserve in southeastern Saskatchewan. Originally from
Maple Creek, he is a graduate of the University of Montana with a specialization in Blackfoot languages, where he “got to know the people and because of my rodeo western background . . . that’s a big deal out there, so there was some common interests and I developed some friendships and they really took care of me.” Rick has been the principal of this school for 11 years.

- Robert Hanson is a first-year principal of a Kindergarten to Grade 8 school situated along the TransCanada Highway in a small rural community near Moose Jaw. He upholds the school’s Learning Improvement Plan that focuses on the reading outcomes of students. Reading as well as an emerging priority with Math achievement is measured by common assessments provided by Tiger Lily School Division and assessments created by staff.

- Sr. Margaret Westlake is Mother Superior and principal of Poitiers Academy, an independent historical high school in southern Saskatchewan. Poitiers provides education for young women and men as day students from Grade 7 through 12. The residence is for young women in Grades 10-12. “Our school recognizes each person as a gift of God with great potential.” Sr. Margaret affirms. “Our specific goal is to give not only a sound academic training but, a complete human and Christian formation.”

5.34 Interviews and instrumentation.

Because this is a study of evidence-based problem-solving for instructional leadership, an initial set of guiding questions was created which focused on: statistical conception and calculation, categorization and inference, effects and consequences, administrative position and causal agency. First draft questions were pilot-tested with two experienced, practising Regina school principals in spring 2012 in tape-recorded interviews in their offices. Based on their feedback, I modified and narrowed the range of questions, and narrowed the objectives for each interview to provide a focal point.

This preliminary set of semi-structured questions was cross-tabulated with my research questions, and with the conjectures offered in my four abductions (see Appendix E). The questions were also ordered to progressively narrow and provide more focus over the three interview sessions, each session building on the insights offered in the previous interview. Each set of questions along with the appropriate Data Display was
e-mailed to each participant at least one week ahead of time for their general orientation prior to meeting. As outlined in Table 5.1, each interview was approximately one and a half hours in length, conducted during the school day, and typically held in the principals’ office or an adjacent meeting room. Two interviews were conducted in a neutral office at the University of Regina, and one was held in a restaurant over supper.

My objectives varied by session: in the first interview, my objective was to gain an overall sense of the principals’ beliefs about the average in relation to their views of leadership purpose(s). In the second interview, my objective was to gain a sense of principals’ ideas about the computational forms of the average they considered and detail about principals’ views of the consequences of averaging. In the third interview, I probed for greater detail using transcripts from interview 1 and 2 to inquire about the principal’s role in relation to computations, consequences, effects, and decision-making with an average. Invariably, the actual interview questions differed markedly in each session as I probed the ideas that principals expressed. I adopted Kvale and Brinkman’s (2009) guidelines for conducting phenomenological interviews, where the interviewer is
(a) centred on the interviewee's life-world, (b) in quest of the meaning of phenomena in his/her life-world, (c) focused on qualitative matters more so than quantities, (d) specific, (e) presupposition-less, (f) attendant to particular themes which emerge, (g) open for ambiguities to be expressed, (h) conversational, and (i) intent on making the interview a positive experience for both. At one or more points in each interview, I played devil’s advocate but indicated that I was about to do so.
Table 5.1
Interview Schedule and Content, Ten Saskatchewan School Principals,
December 2012 – April 2013

<table>
<thead>
<tr>
<th>Interview*</th>
<th>Interview Objective</th>
<th>Interview Tasks</th>
<th>Instrumentation**</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview 1 (Dec. 2012)</td>
<td>To gain an overall sense of the principals’ beliefs about the average in relation to their views of leadership purpose(s)</td>
<td>Complete consent form Introductions Interview 1 questioning</td>
<td>Semi-structured questions</td>
</tr>
<tr>
<td>Interview 2 (Jan.-Feb. 2013)</td>
<td>To gain a sense of principals’ ideas about the computational forms of the average they consider, and detail in terms of principals’ views of consequences</td>
<td>Confirmation and clarification of Transcript 1 Read aloud Data Displays 1 and 2 with follow-up probes Interview 2 questioning</td>
<td>Semi-structured questions Data Display 1 with elementary/secondary alternatives Data Display 2*** Transcript 1</td>
</tr>
<tr>
<td>Interview 3 (Mar.-Apr. 2013)</td>
<td>To inquire about the principal’s role in relation to computations, consequences, effects, and decision-making</td>
<td>Confirmation and clarification of comments on Transcript 2 Read aloud Data Display 3 with follow-up probes Interview 3 questioning Photographs of gestures****</td>
<td>Semi-structured interview questions Data Display 3 with elementary/secondary alternatives Transcript 2</td>
</tr>
</tbody>
</table>

* 1.5 hours duration in principals’ office  
** E-mailed to principals one week before the interview  
*** Used in both Phase 1 and Phase 2 of the study  
**** All photographs taken with respondents’ permission, as replicating gestures made in earlier interviews

All interviews were tape-recorded and transcribed according to the rules suggested by Ives (1980) by a third party, an experienced transcriber who is an employee of the University of Regina. I made and dated field notes immediately after each interview to record my detailed observations. In addition, I verbally summarized my understandings at appropriate times during the interviews, and invited the interviewee’s
(dis)confirmation of my understandings. Transcripts were sent to interviewees for confirmation, and they were invited to make adjustments, or delete segments as inappropriate, if they so wished. A 15-minute segment from one interview was excised as confidential and inappropriate commentary on central office policy. All other interview transcripts remained intact after review by the respondents. In addition, the pertinent content in all tables of data presented in Chapter 6 was confirmed as accurate by each of the 10 respondents.

The focal points in the second and third interviews were three data displays, chosen to reflect different computations and presentations of the average. My criteria when constructing or selecting these data displays were variety in the average by: (a) data source (professional in classroom, Ministry large-scale assessment, commercial tests); (b) calculation (raw scores, scaled scores, criterion-referenced scores); (c) format (numeric, graphic, prose); and (d) level of aggregation (student, classroom, school, school division, provincial, national). However, I maintained some degree of consistency in the objects that were represented in these data displays: K-12 student achievement in key subject areas. I heeded several phenomenologists’ strictures to enable imaginative variation (Giorgi, 1992, 1997; Moustakas, 1994; Polkinghorne, 1989), but in the reader’s circumstance rather than simply during my own subsequent and reductive analyses. Because Peirce defined conception as the reduction of a manifold to sensuous unity (EP1. 1), and because school principals’ conceptions of the average are the focal point of this investigation, variety in data displays as representations and consistency in student performance as object seemed appropriate.

I constructed Data Display 1, with elementary and secondary school variants, based on my own experience in school administration, as a list of average scores for
fictitious individual students. Here the comparators within the data set were endogenous within a particular classroom, and for students’ performance in four key subject areas using a straightforward arithmetic mean. Some fictitious individual and family background attributes were also provided (see Appendix F).

Display 2 was used in both the qualitative and quantitative phases of this inquiry. Two alternate sets of four frames had been previously and provisionally constructed. One set of frames was based in Fuzzy Trace theory, which revolves around memory and mathematics; the other set was based in Prospect theory that revolves around the role of future events and statistical understanding. I eventually chose the frames based on Prospect theory, because commentary during the first interview suggested this would be the most productive approach. With this decision made, I chose one of the four Prospect theoretic frames, which focused on the public release of negatively-construed information, for principals to read aloud in the second interview. Data Display 2 was designed to depict the school average as a percentage of the school population meeting various criterion thresholds for an un-named subject discipline. The display was framed within a school planning process. It did not indicate whether the test scores were generated by teachers through a teacher-constructed test, or through an externally-administered assessment. The normative comparator is with a provincial distribution. Data Display 2 asked school principals to forecast the significant consequences from test outcomes within a planning process for which the school principal has been assigned responsibility. Like Data Display 1, it was deliberately constructed to create administrative doubt (see Appendix F) (Irwin, Levin, Schneider & Gaeth, 1998).

Data Display 3 also came with options, even though all were in the form of histograms that invited normative comparisons. For principals of predominately elementary schools, I drew a presentation intact from a recent study commissioned by the United States Department of Education on American teachers’ data interpretation skills.
Data Display 3 depicts averages in terms of national averages (Means, Chen, DeBarger, & Padilla, 2011, p. 29); the scales for representing these averages are not percentage averages, but rather are scale scores normed to 500, similar to OECD and national assessments. For this display, data are broken out at the school and skill level for a particular grade; the comparison point for an average within the display is longitudinal.

For those principals in secondary schools, I extracted two figures from the most recent provincial education indicators report (Saskatchewan Ministry of Education, 2011). One chart depicts average student marks at the provincial level for key subject areas, and the other represents the same marks in terms of Aboriginal and non-Aboriginal students, male-female students, and rural-urban populations (see Appendix F).

I asked each principal to read the respective data displays in a read-aloud protocol, so that I could understand their reasoning patterns. Because this was a phenomenological and not a clinical interview, particular questions for each Data Display were not pre-scripted, although I introduced each script by asking them to read it aloud as they would do naturally. My field notes would suggest that in about 20 percent of the interviews, it was evident the principal had pre-scanned the material in some detail before the session. Interviews were recorded using a Sony handheld recorder, and an additional backup tape recording was made in case of technical problems. I took photographs with the respondents’ permission; any identifying information was air-brushed or trimmed from these photographs. In addition, I made extensive written field notes immediately after many but not all interviews, depending on the travel time available between meetings. I e-mailed transcripts of interview testimony to participants after each interview, for confirmation of their accuracy in the succeeding interview. In several cases, I asked additional follow-up questions in writing to clarify school principals’ terminology and ideas before proceeding to the second interview, and several principals provided written responses in relation to my semi-structured questions. After the third
interview session, transcripts were emailed with a written request for confirmation of accuracy. In addition, an initial draft of Chapter 6 was sent to respondents for their comments and confirmation that my interpretations of their own interview transcripts accorded with their views.

5.35 Coding and analyses.

Qualitative analyses of tape recorded interviews involved both open and closed coding in four domains. Given this study’s purposes, (to ascertain the inferential habits of school principals relating to the interpretation or reading of averages), I constantly compared school principals’ (a) classifications for both the average as Sign and its object, student achievement, (b) patterns of reasoning in relation to three types of graphic displays for averages, (c) school principals’ testimony about consequences and antecedents in relation to a depicted average, and (d) conceptions of causal sequences and agentic influence in relation to the average.

I did all the open coding myself directly and manually using procedures described in Strauss and Corbin (2008). I chose this approach rather than using a software package and algorithm, because I was well trained in this skill during my previous graduate studies. Another experienced qualitative researcher, a former Professor of Curriculum Studies and Associate Dean of Graduate Studies at the University of Saskatchewan, accomplished an independent verification of my coding. Similar categories resulted from the co-coding by this auditor, but some labeling of the categories differed. Through discussion, we achieved a consensus for labelling the categories.

I am trained as a Canadian school and public administrator and have been closely involved with assessment and evaluation activity in Saskatchewan and across Canada. Therefore, I used a number of procedures to ensure that my familiarity with quantitative processes did not cause me to misread interview transcripts. I sent each transcript back to
the interview participant with follow-up questions in the margins asking for clarification of certain points and each participant had an opportunity to verify the accuracy of the transcript or make corrections. During the interviews, I frequently summarized the interview participant’s comments and asked for feedback on the accuracy of my summary. Although Peirce was explicit in saying that it will be impossible for the phenomenological inquirer to separate his conceptions for a Sign from those of others in communication, I made some effort in the introductory chapter of this dissertation to distinguish my meanings for an average from those of participants. I also endeavoured to set aside my past experience with assessment and evaluation in classroom, large-scale assessment, and policy circles, but those experiences inescapably coloured my questions and analyses.

For the purposes of this dissertation, and consistent with Peirce’s emphasis on classification rather than case depiction, I developed findings as descriptive comparisons and contrasts among the 10 participants, rather than as complete and interpretive case studies. My approach was similar to the constant comparative method recommended by Strauss and Corbin (2008) who emphasize that this method conforms to pragmatic assumptions (pp. 1-8). One researcher has affirmed that constant comparative techniques reflect Peirce’s method of discovery (Reichertz, 2010) and others affirm that grounded theory methods are compatible with research into managerial practices (Locke, 2001; Teddlie, 2005) with strong roots in American pragmatism.

One principle I upheld throughout all data analyses was that school principals’ testimony should be treated as credible. Peirce was insistent that there “is no surer mark of inexperience in dealing with witnesses than a tendency to believe they are falsifying, without any definite, objective and strong reason for the suspicion” (EP2. 113). Another principle was that coding of categories must precede coding of inferences. Accordingly,
documentation from all three interviews was considered when open-coding to categories were assigned to classifications, consequences/antecedents, and causal/agentic influences. In contrast, those inferential patterns that were closed-coded came only from the second and third interview transcripts, which featured commentary offered during and shortly after actual reading of the three data displays, not through the general introductory questioning (Bogdan & Biklen, 1998; Marshall & Rossman, 1989) during the first interview session.

Therefore, I proceeded to review and analyze the transcripts for their substantive content, with six questions in mind: What is the category? When does the categorization occur? Where does the categorization occur? Why does the categorization occur? How does the category arise? With what consequences does the categorization happen? I made extensive marginal notes in relation to these questions, using a highlighter to underline key passages. I eventually numbered the highlighted sections to determine their relative emphasis across the full set of 30 interviews, assigning each key passage one point. For this study’s purposes, the relevance of each categorical pattern and hence causal relation within each interview was determined by using a one-half ratio threshold. That is, the total number of times an idea was coded was divided by the total number of respondents (i.e., over half of the 10 respondents made the same point in one or several of the interviews).

To precisely ascertain inferential patterns of reasoning, I reviewed and recoded again only highlighted sections according to the rules in Peirce’s logic as applied to documentary evidence, and as especially pertinent to testimony (EP2. 75-114). This involved the difficult task of identifying respondents’ premises and conclusions with any given average(s), which I noted directly in transcripts. Distinctions between Peirce’s two types of deduction, three types of induction, and six types of abduction— as described by
Kehle & Cunningham (2000)—were discernible but often intermixed within a reading or interpretation of a single data display. Where judgements were required, I used the same one-half ratio rule (i.e., over half of the inferences made by an individual school principal in relation to one data display fell into a particular abductive, deductive, or inductive mode of reasoning). Where interview time constraints or interruptions (school bells or visitors at the door) meant that the principal did not have what I considered to be fair opportunity to respond to the data display, I registered the reasoning pattern as 0 and a blank in the EXCEL software in which I recorded overall patterns.

Two spring 2013 debriefing meetings with principals and vice-principals in one school division who had participated in the quantitative phase of the study were a helpful opportunity to gauge the plausibility of the findings from the qualitative phase. A further June 2012 meeting with a 17-member group taking a University of Regina graduate class in educational administration provided another forum to gather feedback and clarify preliminary findings from the qualitative phase of my research.

5.4 Quantitative Method for Phase 2

5.41 Assumptions.

A theory of statistical inference was developed by Charles S. Peirce in a series of papers entitled *Illustrations of the Logic of Science* (EP1. 109-224) and in his *A Theory of Probable Inference* (1883). In another publication, Peirce emphasized the importance of randomization-based inference in statistics. Peirce randomly assigned volunteers within a blinded, repeated-measures design to evaluate their ability to discriminate weights; he asked them to say whether they sensed a balance beam weight was lighter or heavier, with pre- and post-measures (Peirce & Jastrow, 1885). His experiment (con)tested Fechner’s findings in the psychophysics of perception. Such early research into the intensity of sensation undergirds Kahneman’s and Tversky’s approaches in Prospect
theory (1982) to account for the role of perception on decision-making (Tversky & Kahneman, 1974, 1986, 1992). Peirce's experiments in the early 1880s inspired subsequent researchers in psychology and education, who developed a tradition of randomized experiments in the late 1890s (Hacking, 1988), distinct but not substantively different from the theories of randomization in Fisher’s studies of agricultural plots in the 1920s (Dehue, 1997; Hacking, 1988; Hall, 2007).

Cognitive science experiments in perception often draw on the psychophysics of perception when studying what is known as framing effects. Framing research involves changing the perspective brought to information, to ascertain whether certain conditions in the environment influence perception (Levin, Schneider, & Gaeth, 1998). Two kinds of frames may be used. Cognitive psychology often relies on “equivalence frames” to determine whether logically-identical information influences thought patterns (i.e., the odds are 50-50 the patient will survive; 60 of the 120 patients will die). “Emphasis frames” are often considered in media and marketing studies; they simplify a situation or issue by narrowing or focusing on a subset of attributes, ignoring others to encourage desired interpretations and discourage others (cornflakes have 75 percent healthy ingredients/cornflakes have 25 percent unhealthy ingredients). This phase of the study merges both types (Sher & Mckenzie, 2011).

The effects of such frames on school administrators’ planning and problem-solving approaches were identified through a 30-minute inventory of their cognitive styles or dispositions. This quantitative phase looks at the anticipated effects of below- and above-average student performance on principals’ tendency to become more or less rational-analytic or intuitive-experiential in their dispositions. Whereas in the qualitative phase of this study, the principal was bound by my questions in her or his office during three interviews, in the quantitative phase, the administrators’ environment was
controlled through his or her assignment to one of four groups, where one of four
pre-determined frames was randomly assigned for interpretation. The experiment brings
an explicitly normative outlook to sampling, instrumentation, and analysis. That is, the
statistician-as-author’s logic was superimposed from without in terms of an *ad hoc* norm.
Notions of possibility were implicit in the qualitative phase, which revolves around the
practical significance of an average, but in the second quantitative phase, significance is
probabilistically ascertained through the calculation of effect sizes.

5.42 **Experiment design and decisions.**

Initial findings from the first round of interviews in the qualitative phase informed
the selection of frames for the second and, to a certain extent, overlapping phase of the
inquiry. A preliminary design decision was whether Fuzzy Trace theory or Prospect
type better reflected school principals’ quantitative reasoning in relation to average
student achievement. Fuzzy Trace theory postulates that adults generally and
retrospectively reason with vague, qualitative gists of numeric information that are stored
in long-term memory, but that adults accomplish arithmetic tasks in verbatim or working
memory. Fuzzy Trace theory assumes that denominator neglect and class inclusion
difficulties explain many cognitive distortions with ratio concepts: the interaction of these
two memorial operations explain many difficulties that adults have in judging risk. In
contrast, Prospect theory postulates that adults make evaluative judgements through
quantitative weightings of risk for impending or anticipated action. That is, adults reason
in terms of gains and losses from situationally-defined base points in uncertain situations
and not in terms of utility. Both theories assume that an adult’s judgements may be
distorted by a variety of perceptual biases when outwardly scanning the environment.

Based on field notes and initial transcripts from my first round interviews, I
decided that Prospect theory better accorded with principals’ testimony. Interviews
conducted during the phenomenological phase revealed that: (a) most school principals
had very detailed and sharply-defined, long-term memories of some statistics relating to their own achievements as students and the achievements of others, along with the issues relating to those statistics; (b) school principals were intensely aware of denominators when averages were discussed; and (c) school principals were more attuned to the prospective or future influence of averages than to their influence in the past. Moreover, a thorough reading of Reyna and Brainerd’s writings (Reyna, 2012; Reyna & Brainerd, 1990, 1991, 1995a, 1995b, 2011; Reyna, Lloyd & Brainerd, 2003; Reyna, Nelson, Han & Dieckmann, 2009), and two subsequent discussions with a University of Regina developmental psychologist with expertise in false memory persuaded me that Fuzzy Trace theory as a set of psychological presuppositions is logically inconsistent with its purported mathematical foundation in Fuzzy set logic. Fuzzy Trace theory fails to precisely demarcate temporally between long-term and working memory, removing an important control in a controlled experiment. In addition, a Prospect theory model better resembles Peirce’s orientation to interpretation in many but not all respects: it assumes that school principals as readers of statistics would look forward to anticipated consequences flowing from planning, rather than looking backward to results already received or introspectively to memories already encoded. Prospect theory also enables the researcher to calculate and reason quantitatively under all framing conditions; it does not restrict the researcher to qualitative (or prose) modes under some conditions. Nor does it require “timed tests” of memory, rendering it more feasible for application in an administrative meeting context. On this basis, I chose Prospect theory as the underlying psychological basis for designing four alternate data displays with accompanying instructions.

Each data display framed an identical collection of anticipated student test scores depicted as a distribution within five categories as represented in an histogram (Appendix G) for a school. As common conditions within the task, school principals
were assigned responsibility for preparing a school level plan. They were informed that if their plan were followed, the resulting student scores would be asymmetrically distributed. Each display was multimodal: that is, the averages were depicted in graphic, numeric, and linguistic forms. Such multimodal displays are used extensively to communicate cognitive science research, and are usually used in indicators reports distributed by Statistics Canada and provincial ministries of education across Canada. Each display featured an identical graphic depiction, the same distribution of scores, and only those prose commentaries and numeric categorizations necessary to consider two factors. The two factors were the prospect of (a) publicity (or audience)—that is, student test scores would become public or would remain confidential to the school, and (b) positivity (or valence)—that is, student achievement would be labelled in prose and numeric text positively as above average, or negatively, as below or at average (see Appendix G). These two factors were then cross-combined to yield four different lenses or frames for an identical distribution of scores. Each average was depicted as a percentage of the school’s student population reaching one of five criterion performance standards. Moreover, the percentages were all footnoted as probabilistically significant at less than a .001 threshold in comparison with provincial scores, without indicating whether the focal school scores were lower or higher than the provincial norm.

The instructions asked school administrators to proceed in four steps. First, they were asked to complete the Rational Experiential Inventory that determined whether their problem-solving pre-dispositions were primarily rational or primarily experiential. Second they were asked to supply basic demographic information. Third, they were asked to read one of the randomly-assigned, multimodal data displays. Fourth, they were asked to complete a scrambled version of the same Rational Experiential Inventory measure to determine whether their problem-solving post-dispositions were rational or experiential.
The overall four frames by two measures experimental design can be depicted as four orthogonal cells, bracketed by the pre- and post-measures, illustrated in Figure 5.2 below. Therefore, we have a 2 (pre-post) by 2 (positive-negative) by 2 (public-confidential) by 2 (rational-experiential) design.

<table>
<thead>
<tr>
<th>Pre-disposition measures (Rational-Experiential Inventory)</th>
<th>Frame 1: Positive Confidential</th>
<th>Frame 2: Negative Confidential</th>
<th>Post-disposition measures (Rational-Experiential Inventory)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame 3: Positive Public</td>
<td></td>
<td>Frame 4: Negative Public</td>
<td></td>
</tr>
</tbody>
</table>

*Figure 5.2.* Research design into school administrators’ perceptions of average student achievement, pre- and post-disposition measure and four frames.

In short, the pre-post problem-solving disposition measure became two dependent variables, while the audience and valence were two independent variables. For the purposes of MANCOVA analyses, the frames were between-subjects variables. Dispositions to be either rational-analytic or experiential-intuitive in their problem-solving or cognitive style were considered as within-subject, dependent variables. For the purposes of the regression analysis, the six demographic variables serve as independent variables while the single, summative or combined scale was considered as a dependent variable. Likewise, for the analysis of variance comparing principals’ and vice-principals’ pre-dispositions, this same summative or combined scale served as a dependent variable.

**5.43 School division sampling and randomization.**

Appropriate sample sizes were initially calculated *a priori* using the G-Power software package (Faul, Erdfelder, Lang, & Buchner, 2007), using Cohen’s guidelines (1992, Table 2, p. 158). The n/Ns for medium effect size at power .80 and set at .05 for tests of mean difference and significant *r* are listed as 64 and 85, respectively. While I
ensured the total sample size was larger than 200, actual sample sizes within the four “treatment” frames were limited by time and budget considerations. School administrators included both principals and vice-principals, but not central office school officials. Including vice-principals enabled the overall sample size to be expanded; permitted conformity with the typical format of school division meetings where both principals and vice-principals attend together; recognized that school principals often delegate aspects of their instructional leadership responsibility to vice-principals; and directly enabled testing of the two research hypotheses which deal with the relationship between administrative position and problem-solving disposition.

Five Saskatchewan school divisions gave written permission for their administrators to participate in the project during February and March 2013, while at division-wide administrative meetings. Two participating school divisions were large urban divisions, two were large rural divisions, and one was a small division in a small, urban setting in Saskatchewan. Of these, two were Roman Catholic school divisions. I was present at five of the six meetings, where a senior district official (superintendent or assistant director) briefly introduced the study and I explained its purposes and procedures. As detailed in Table 5.2, the total sample of 210 school principals and vice-principals was randomly assigned into four groups, each similar in its size and general demographic composition. According to Ministry of Education data (Saskatchewan Ministry of Education, 2010, p. 119), this sample represents approximately 18 percent of the total number of principals and vice-principals in the Province (1,182 in 2010). Each group also represented proportionally similar percentages of the provincial administrative population by position (principal/vice-principal) and by gender. Informed consent was obtained from individual participants in all cases. I am cautious in claiming that this sample statistically represents the entire population of Saskatchewan school administrators, since neither First Nations authority school administrators, nor those in the Northern Lights School Division were involved.
The overall profile for the total sample and randomized groups is outlined in Table 5.2 below.

**Table 5.2**

**Randomized Total Sample and Group Composition by Frame, Phase 2 Experiment**

<table>
<thead>
<tr>
<th></th>
<th>Group Frame 1</th>
<th>Group Frame 2</th>
<th>Group Frame 3</th>
<th>Group Frame 4</th>
<th>N</th>
<th>Proportions</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Principal/Vice-Principal</td>
<td>30/20</td>
<td>28/24</td>
<td>34/31</td>
<td>32/20</td>
<td>137/73</td>
<td>65%/35%</td>
</tr>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male/Female</td>
<td>36/14</td>
<td>30/33</td>
<td>25/30</td>
<td>18/34</td>
<td>109/100</td>
<td>52%/48%</td>
</tr>
<tr>
<td><strong>Total Professional Experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6-10 years</td>
<td>3</td>
<td>4</td>
<td>4</td>
<td>2</td>
<td>13</td>
<td>6%</td>
</tr>
<tr>
<td>11-15 years</td>
<td>10</td>
<td>10</td>
<td>9</td>
<td>7</td>
<td>36</td>
<td>17%</td>
</tr>
<tr>
<td>16-20 years</td>
<td>12</td>
<td>12</td>
<td>14</td>
<td>14</td>
<td>52</td>
<td>25%</td>
</tr>
<tr>
<td>&gt; 20 years</td>
<td>26</td>
<td>26</td>
<td>28</td>
<td>29</td>
<td>109</td>
<td>52%</td>
</tr>
<tr>
<td><strong>Administrative Experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt; 1 year</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>4</td>
<td>12</td>
<td>6%</td>
</tr>
<tr>
<td>1-5 years</td>
<td>18</td>
<td>17</td>
<td>17</td>
<td>18</td>
<td>70</td>
<td>33%</td>
</tr>
<tr>
<td>6-10 years</td>
<td>11</td>
<td>14</td>
<td>17</td>
<td>10</td>
<td>52</td>
<td>25%</td>
</tr>
<tr>
<td>11-15 years</td>
<td>11</td>
<td>12</td>
<td>15</td>
<td>11</td>
<td>49</td>
<td>23%</td>
</tr>
<tr>
<td>&gt; 15 years</td>
<td>7</td>
<td>7</td>
<td>4</td>
<td>9</td>
<td>27</td>
<td>13%</td>
</tr>
<tr>
<td><strong>Postsecondary Education</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 years</td>
<td>5</td>
<td>8</td>
<td>4</td>
<td>5</td>
<td>22</td>
<td>11%</td>
</tr>
<tr>
<td>5 years</td>
<td>15</td>
<td>7</td>
<td>11</td>
<td>8</td>
<td>41</td>
<td>20%</td>
</tr>
<tr>
<td>6 years</td>
<td>11</td>
<td>17</td>
<td>15</td>
<td>10</td>
<td>53</td>
<td>25%</td>
</tr>
<tr>
<td>7 years</td>
<td>3</td>
<td>5</td>
<td>11</td>
<td>12</td>
<td>31</td>
<td>15%</td>
</tr>
<tr>
<td>&gt; 7 years</td>
<td>17</td>
<td>15</td>
<td>14</td>
<td>17</td>
<td>63</td>
<td>30%</td>
</tr>
<tr>
<td><strong>Self-Efficacy Statistics</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Very weak</td>
<td>2</td>
<td>0</td>
<td>1</td>
<td>2</td>
<td>5</td>
<td>2%</td>
</tr>
<tr>
<td>Weak</td>
<td>6</td>
<td>5</td>
<td>6</td>
<td>9</td>
<td>26</td>
<td>12%</td>
</tr>
<tr>
<td>Average</td>
<td>25</td>
<td>28</td>
<td>32</td>
<td>29</td>
<td>114</td>
<td>54%</td>
</tr>
<tr>
<td>Strong</td>
<td>12</td>
<td>16</td>
<td>12</td>
<td>10</td>
<td>50</td>
<td>24%</td>
</tr>
<tr>
<td>Very strong</td>
<td>6</td>
<td>3</td>
<td>3</td>
<td>2</td>
<td>14</td>
<td>7%</td>
</tr>
</tbody>
</table>

**Note.** Subcategory proportions may not total 100 percent due to rounding. One consultant who divided central office and school administrative responsibilities was designated as a principal. One participant did not supply gender information.
5.44 Instrumentation.

To gauge school administrators’ problem-solving dispositions in the five Saskatchewan school divisions between February and April 2013, I administered the Rational-Experiential Inventory (REI-40) to both principals and vice-principals, combined under the label of in-school administrators. The 40-item, self-report inventory, as developed by Dr. Seymour Epstein, posits two information processing systems in the human mind—an experiential approach and a rational or logical approach (Epstein, Pacini, Denes-Raj, & Heier, 1996). Inventory questions, such as “I believe in trusting my hunches” and “I hardly ever go wrong when I listen to my deepest gut feelings to find an answer” presume the experiential system is holistic, associated with affect, and operates on the basis of schemas acquired from lived experiences. Intuition is assumed to inhere within the experiential system, and to operate automatically, pre-consciously, nonverbally, rapidly, effortlessly, and concretely. In contrast, REI-40 questions such as “I enjoy solving problems that require hard thinking” and “Using logic usually works well for me in figuring out problems in my life” deal with an inferential logical system that operates consciously, primarily verbally, slowly, and effortfully. According to Epstein, the rational system is abstract, analytic, affect-free, and evolutionarily the more recent of the two systems (Epstein, 1994, 2010) similar to that propounded within Kahneman’s Prospect theory.

Within this model of the mind, the degree to which either system dominates thought and behaviour is a function of: (a) the extent to which the situation is associated with a customary way of responding, (b) the degree of emotional involvement, (c) experiential dominance based on repeated amounts of relevant experience, and (d) “individual differences in preference for relying on one system more than the other” (Epstein, Pacini, Denes-Raj, & Heier, 1996, p. 391). The specific properties of this model of mind (Epstein, 1973, 1994) are summarized in Appendix H.
The REI instrument thus measures both the rational (System 2) and experiential (System 1) disposition within an individual’s cognitive style (Epstein et al., 1996; Pacini & Epstein, 1999; Pacini, Muir, & Epstein, 1998). The rational (originally derived from the Need for Cognition scale developed by Cacioppo & Petty (1982)) and experiential thinking scales are composed of two subscales each, Rational/Experiential Ability and Rational/Experiential Engagement. The 40 items are evenly distributed among the four subscales. Participants evaluate each item and indicate how true each belief statement is about oneself on a 5-point scale ranging from 1 (completely false) to 5 (completely true). The scores obtained are derived from simple sums of the two respective subscales when considered in combination. The higher the score, the higher is the disposition within that thinking dimension.

The REI-40 has been extensively validated concurrently, construct-wise, and criterion-wise, to demonstrate its accuracy and trustworthiness in the English language across North America, and in translation to other languages as well (Bjorklund & Backstrom, 2008; Gunnell & Ceci, 2010; Hodgkinson, Sadler-Smith, Sinclair, & Ashkanasy, 2009; Marks, O’Neill, & Hine, 2008; Novak & Hoffman, 2009; Witteman, van den Berken, Claes, & Godoy, 2009; Wong, Kwong, & Ng, 2008). Furthermore, the instrument has proven sufficiently robust to register changes in decision styles across several psychological theories. Previous criticism of the REI has been for its use as an independent variable for predicting framing effects on memory (Reyna & Brainerd, 2008), not as a dependent variable for demonstrating the effect of perception on decision-making or problem-solving styles. A French version of the REI questionnaire has been concurrently validated and used by Stadelhofen, Rossier, Rigozzi, Zimmermann, and Bethoud (2004) in a cross-cultural context. The correlation between rational and experiential thinking in that study was $r (490)=0.07$ (n.s.), which is consistent with other studies (Epstein, 2003) and supports the parallel nature of the two thought systems.
To establish the REI-40’s construct validity in the Saskatchewan setting, I asked two principals during a December 2012 pilot test of my questions, and four central office administrators during the December 2012-January 2013 pilot of the on-line version of the experiment, to review and affirm that the instrument was sufficiently generic and appropriate for efficiently gathering data consistent with this study's purposes. Both principals and three of the four superintendents responded. No substantive criticisms were offered, other than that two REI-40 items might be slightly adjusted. Two REI statements, which inquired about the use of “logic” and use of “gut feelings” in “figuring out problems in my life, were accordingly rephrased to “figuring out problems in my administrative practice.”

In addition, but not reported in detail here, I conducted a principal components analysis (PCA) post hoc with the pre-disposition measure’s 40 items with SPSS18—to affirm that the instrument was tapping two, distinct, dependant variables—rational-analytic and experiential-intuitive dispositions. Inspection of the correlation matrix revealed many coefficients of .3 and above. The Kaiser-Meyer-Olkin value for sampling adequacy was .83, exceeding the recommended value of 6 (Kaiser, 1970, 1974). Bartlett’s Test of Sphericity (Bartlett, 1954) reached statistical significance (.000). Principal Components Analysis revealed the presence of 10 components with eight values exceeding 1, explaining 17.4, 16.5, 5.2, 3.9, 3.6, 3.4, 3.2, 2.9, 2.7, and 2.5 percent of the variance respectively. The scree plot, considered in terms of Catell’s (1966) test, revealed a clear break after the second component. A visual review of component matrix content also supported a two component solution, which accounted for 33.9 percent of the total variance within the instrument. No parallel analysis was conducted.

Given this study’s overall exploratory goals—to ascertain the applicability of Peirce’s theory of interpretation to leadership practice, and not to develop new theories of
an integrative personality, nor to confirm models of perception with novel instrumentation, it therefore seemed both reasonable and economical to rely on existing measurement instruments. Using the instrument as a dependent variable enables me to test Peirce’s syncretic theory, through a study of interactions and continua—and its counterfactual—as separate and independently operating systems, as upheld in Prospect theory. Moreover, the REI-40 meets several criteria for validity. The Inventory is predicated on statements of belief, as is Peirce’s theory of interpretation. The instrument taps problem-solving dispositions rather than decision-making criteria, thereby matching Peirce’s view of the reader. Furthermore, the instrument presumes an integrative theory of the personality (Epstein, 2003), as did Peirce (EP2. 330-331). The REI-40 registers changes in problem-solving styles across a scale as a continuum rather than in terms of dichotomous reading tasks; continua are foundational to Peirce’s synechism and his pragmatism (EP2. 2-3; EP2. 312-324; RLT. 242-267).

Furthermore, the instrument’s constructs uphold many of the same assumptions about an adult decision maker as Peirce upheld: that personal and formal scientific theories of reality are similar in their properties; that ”intuitions” arise in pre-consciousness, not the unconscious (as percepts in Firstness); that humans are animated by fundamental underlying needs or instincts; that everyone develops a personal theory of self and the world as a set of propositions that work in symbiosis; and that “personal theories of reality, in common with scientific theories serve the purpose of organizing the data of experience and directing behaviour” (Epstein, 1991, p.118). These propositions are compatible with both social psychological views (Strack & Deutsch, 2006) of reflective and impulsive behaviour, and with individual differences in information processing (Stanovich & West, 2000). Therefore, Epstein’s model of the mind is compatible with a view of personality as a coordination of ideas (EP1. 331) or bundle of
habits (CP6. 228) in a social setting that is animated by purpose, as in Peirce’s formulation. As Peirce said, “pragmatism and personality are more or less from the same kidney” (EP2. 134).

Nevertheless, there are several limitations in the REI relative to this study’s purposes. If we define intuitive modes of thinking as a “fast and frugal” decision heuristic or as behaviour, then neither speed of thinking nor actual behaviour are directly measured. Nor does the REI-40 instrument, as an amalgam of several scales in cognition, provide especially detailed information about particular elements in the experiential domain—creativity, imagination, and the like (Handley, Newstead, & Wright, 2000), raising questions about which aspects of experience are actually being measured. A newer version of the REI taps these constructs more fully, but at the expense of yielding detailed information in the rational-analytic dimensions. Similar issues are raised in the logical-analytical domain, because it does not assess various forms of reasoning. After weighing these deficiencies, I decided that the advantages of the REI-40 outweighed its shortcomings: weaknesses in the REI-40 were offset by extending the interviews conducted during the qualitative phase of this study to gather potentially contradictory or complementary data that would shed light on heuristic aspects of thinking and its impact on belief.

5.45 Data collection and entry.

Quantitative data were collected in six school division meetings between February and April 2013, using either electronic or paper format, with assistance from the Social Sciences Laboratory (SSRL) at the University of Saskatchewan’s College of Arts and Science in Saskatoon. The Rational Experiential Inventory-40 was loaded twice—as a pre-disposition measure completed before participants had read one of four different multimodal data displays (frames) and as a post-disposition measure completed after
participants had read the data display—on specially-designed, secure, website software. Six background demographic variables and one of the four frame panels were randomly inserted between the pre- and post-disposition measures to provide sufficient time for 30-second working memory to lapse. For the post-disposition measure, inventory questions were scrambled and the computer software was syntactically-programmed to prevent respondents from revisiting and thus revising their pre-disposition responses after reading the data display, or from simply copying pre-disposition responses in the post-disposition inventory. Scrambling the questions in the paper version also ensured a reliable measure of post-disposition; my visual scan revealed no evidence that respondents revised their pre-disposition responses after reading the data display. The SSRL software was configured to randomly assign one of the four alternate frames in order of log-in time. A confidentially-assigned participant identifier was distributed randomly at each of five school division office sites on a piece of paper with log-in instructions. Thus, software automatically recorded respondents’ responses in a University of Saskatchewan database after they had logged on, but maintained their anonymity.

Pilot testing of all on-line procedures and materials was accomplished by inviting five superintendents in participating school divisions to log on and complete the entire questionnaire during December 2012 and January 2013. Detailed commentary for revising instructions and sequences was received from three superintendents, along with their own REI responses. Based on this advice, modifications were made to procedures and clarifications were made in early February 2013 to the instructions and framing materials themselves. All pilot data were erased from the database in the first week of February before actual administration began.

During the “live” administration between February and April 2013, respondents were given the option of doing the inventory in either electronic or paper form at three
sites. At the other two sites—one rural and one urban—Internet connectivity issues necessitated completion of the paper version. Individuals completing electronic versions typically used handheld devices or computer notebooks at the meeting room table. At survey sites, I personally distributed paper versions from a pile that had been pre-scrambled to ensure randomization. The principals and vice-principals participating were asked to complete the survey autonomously and not to speak with colleagues. Approximately 70 surveys were completed electronically and 140 surveys were completed in paper form. Most respondents took approximately 30 minutes total to finish the entire experiment,\footnote{In this experiment, the major controls were the summative pre-disposition measure, the consistency in task instructions, and the five, similar, central office-sponsored contexts for meetings of principals and vice-principals from across each division, with central office administrators. The summative pre-disposition measure was covariate in (M)ANCOVA analyses, but dependent variable in multiple regression and ANOVA analyses for demographic and principal/vice-principal comparisons.} which included completing the pre-disposition REI measure, supplying information for the six demographic variables, interpreting the randomly-assigned frame, and completing a scrambled version of the REI post-disposition measure.

Manual data entry of paper versions was supervised and accomplished by the Survey and Group Analysis Laboratory manager and one University of Saskatchewan graduate student at the Social Sciences Research Laboratory in May 2013. Data entry began with the graduate student entering each individual paper survey into the database using the original website as entry point, and the participant’s confidentially-assigned number which had been stamped on the cover page of each paper version. The identifying number was used to match an individual respondent with their responses and to ensure that appropriate responses were entered for each survey: no personal information was available to identify participants. The manager thereafter entered the relevant codes for the specific frame used in each paper version, while also scanning data previously entered by the graduate student for inconsistencies. This process included
(a) inspection of the data set for mistakes that may have occurred during manual entry, and (b) checking that the skip logic in software syntax was correctly programmed and questions were correctly registering in the database while the web's electronic version was used by the graduate student. Once satisfied, the manager then proceeded to double-check visually a large subsample of entries made by the student to ensure that the database record was congruent with the actual paper version. Finally, the manager double-checked that the correct frame panel was accurately coded in the database, by comparing each paper copy with the electronic data set, based on the individual survey’s unique identifier. Once the Social Sciences Laboratory manager verified all data in the set, he stripped all respondents’ confidentially-assigned numbers from the electronic data file to ensure anonymity, and transferred to me in mid-May 2013 an EXCEL file which included data from both paper and electronic versions.

5.46 Data analysis techniques.

For this study, to control for the potential impact of school administrators’ prior experience in completing the Rationale Experiential Inventory as pretest, I considered the pre-disposition measure as a covariate in an initial MANCOVA analyses. The aim was to reduce the probability of a Type II error (whenever a false null hypothesis is not rejected) when main or interaction effects are tested, or when comparisons are made within planned investigations. Because the probability of a Type II error is inversely related to statistical power, an ANCOVA is more powerful than its ANOVA counterpart, presuming that other variables are held constant. The $F$-tests associated with a standard ANOVA are computed by dividing the $MS$ for error into the $MS$s for main and interaction effects. If these error terms can somehow be made smaller, then the calculated $Fs$ are larger, $ps$ are smaller, and there is a better chance that null hypotheses will be rejected. When a good covariate is used within such analyses, this is exactly what happens.
To prepare for such analyses, all reversely-worded items were re-coded in analyses to preserve ordinality. For the ANOVA and multiple regression analyses, I used a summative scale that combined the various subscales in the Rational Experiential Inventory into a single measure as dependent variable, ranging from 1 through 200. I considered 100 as the threshold for predominately rational/analytic, set at 200 (high) and 1 (low) set as highly experiential-intuitive. Therefore, the summative scale necessitated further, reverse-coding of experiential-intuitive items. For MANCOVA analyses that contrasted the two different modes of thinking with conditions, ordinality was preserved from high to low within each domain.

All analyses are based on the arithmetic mean with no weightings in either numerators or denominators. With the exception of one analysis where one participant’s gender information was not provided, all means were calculated with a denominator of 210 overall, or the Group Frame N reported in Table 5.2, as appropriate. All rounding was half away from zero (or towards infinity, as consistent with Peirce’s logic (EP1. 319-320) and programmed in SPSS syntax, namely: if the fraction of y is exactly 0.5, then \( q = y + 0.5 \) if \( y \) is positive, and \( q = y - 0.5 \) if \( y \) is negative. For example, 27.546 was rounded to 27.55, and \(-23.735\) was rounded to \(-23.74\), thereby introducing slight positive and negative biases. Two incomplete surveys were removed at the point of data entry, and two other surveys were removed at the point of data analyses because large portions of the post-disposition measure were not completed. There were a total of 210 usable surveys. Within the entire usable collection of surveys, there were 56 missing values out of a total 16,800 REI inventory items. After conducting Little’s MCAR analyses, which showed no systematic pattern for missing variables, I used Estimation Maximization in the SPSS (AMOS 18) facilities to impute missing values to complete the data set. My initial analyses revealed that the instrument was reliable: the Cronbach’s alpha based on
standardized items for rationality was .87, for experientiality was .88, considered separately. The overall alpha for the instrument was .85 as a pre-disposition measure.

Of course, an overall presumption is that the four “treatment” frames are cast in sufficient extremes (80 percent negative/20 percent positive (valence); confidential within school/in the public domain (audience)) that they alone were operative as independent variables, and did not compete with other contextual influences. Instruments were administered during the same quarter of the year within an overall school administrative cycle, in all cases preceding annual budget announcements for the succeeding school year. Moreover, I also presumed that pencil and paper and online collections were sufficiently consistent to not undermine reliability. Data obtained through the online submission to the University of Saskatchewan’s Social Science Research Laboratory (n=68) were collected in identical conditions with those submitted in pencil and paper form (n=146), the major differences being the availability of Internet connections, laptop or handheld technology, and comfort of the participants in one or the other medium. I also presumed that randomizations within each of the five participating school divisions could be combined in a total randomized sample pool, because the requisite instructions, settings in the context of division-wide meetings, and attending personnel were all consistent. Moreover, all statistical assumptions held in analyses when tested within and among the four orthogonal quadrants, including: normality, linear relationships between pre-disposition and post-disposition measures, and homogeneity of variances and regression slopes.

5.5 Presentation of Administrators’ Conceptions of the Average

Phenomenological research methods are often used to gather data about respondents’ experiences with the situation of interest. As Peirce noted, the various categories that people hold in their beliefs, their interpretations of a Sign in relation to an
object, and their beliefs about the immediate consequents associated with a sign are all related to purposes. Phenomenological research methods permit direct but unavoidably partial access to respondents’ beliefs and habits with an average. The major disadvantage is that such methods do not reveal the consequents of the sign except as a conditional possibility.

In contrast, to understand the effect of the average sign as a conditional probability, controlled experiments enable the researcher to hold constant the major features of interest and to register effects with well-validated instruments. Peirce was a pioneer in the use of randomized experiments, and had developed procedures for numerically testing the effects of such signs for making predictions. MANCOVA procedures have been refined over the past century to control for the impact of instrumentation and for prior beliefs when accomplishing such experiments. The primary difficulty with these research designs is that they do not directly reveal principals’ actual conception of the average, but only indirectly their effects of reading an average in relation to alternate lenses brought to the interpretation of a statistic.

Accordingly, the next two chapters present the school principals’ conceptions of the average from two vantage points. Chapter 6 describes the composition of the average as possible consequents. That is, the chapter describes how various forms of the average are manifest in the verbal testimony of school principals as readers when describing their daily practice, and when actually interpreting averages. Thus, it describes what school administrators think in practice as substantive belief-habits. Chapter 7 describes the average, when framed with different conditions, from the point of view of a statistician-author in terms of probable effects on administrators’ problem-solving dispositions as beliefs. It reports on the statistical significance of an average’s effect on school administrators’ mindsets, emphasizing how beliefs change from the point of view of
psychological mechanisms. Within both chapters, findings are presented in chart form, and are thus unavoidably reductive exercises, so we should not presume exhaustiveness of the range of potential meanings. Both chapters represent the findings only insofar as was apparent during the 2012-2013 school year at multiple points in time.

5.6 Recapitulation

Methodologically-speaking, this inquiry is founded on the tenets of classical pragmatism. Hence:

1. An average can simultaneously be conceived as a statement of quality and of quantity. The average may also be studied as a proposition about conditional possibility in terms of consequents, and as an assertion of conditional probability in terms of effects.

2. Qualitative information about 10 Saskatchewan school principals’ conceptions of average student achievement were collected in the first phase, according to the phenomenological precepts in Peirce’s writings.

3. Tape-recorded, phenomenological interviews were conducted with 10 school principals on three different occasions between December 2012 and April 2013. School principals read alternate data displays that depicted an average from several vantage points.

4. Interview questions focused on school principals’ categorizations in relation to the average as Sign, the patterns of reasoning associated with the average, and their causal/agentic ideas with the average.

5. Transcribed interviews were coded using Grounded Theory approaches and constant comparative techniques with both open- and closed-coding to conform with C.S. Peirce’s original writings.

6. Findings from this phase led me to discount theories that suppose averages are processed in the memory, or that school principals may provide false testimony.

7. Quantitative information about the effects of framing an average on 210 Saskatchewan principals’ and vice-principals’ cognitive styles were collected in February-March 2013 within a second phase of this study, following procedures that Peirce pioneered for randomized studies of perception.

8. School administrators were randomly assigned in a controlled experiment to one of four different groups and asked to read an average that was depicted within an asymmetric distribution. Formally equivalent averages were framed according to positive and negative features within the display, and to whether the average would remain confidential to teachers within the school or would enter the public domain.
9. MANCOVA and multiple regression techniques—as 20\textsuperscript{th} century refinements to Peirce’s theory of probable inference—were used in a multivariate design to test framed factors which would influence school administrators’ conceptions of average student achievement in terms of effects.

10. This inquiry thus contemplates both the practical and statistical significance of average student achievement for Saskatchewan school administrators. Significance, or the meaning for a sign, is central to Peirce’s pragmatism.
Order is not pressure which is imposed on society from without, but an equilibrium which is set up from within.

Jose Ortega y Gasset

Chapter 6: Ten Principals’ Conceptions of Average Student Achievement

In this chapter, 10 Saskatchewan school principals’ ideas about the average are presented in terms of their conceptions of and reasoning patterns with student achievement reports. Their conceptions for the average as index and symbol, their gestures and spatial awareness, and their views about the consequents and effects of an average are highlighted within the assumptions of Peirce’s phenomenology.

6.1 Conceptions and Calculations

The meaning that any individual assigns to an average depends on that individual’s perspective. School principals interpret averages from their position as the embodiment of conventional community values and as a representative of the education establishment. When reading statistics, each school principal emphasizes certain elements of an average over others; thus principals’ conceptions of the average are multiple. What an average signifies may, therefore, be understood only in relation to principals’ purposes, the experiences they bring to the number for potential action, and the variety of mathematical forms through which they ascertain the middle.

In Louise’s words, “…it’s a concept we don’t think much about, we just do,”10 echoing Peirce about the automaticity of habit (EP1. 201-202). So too does this principal affirm that the average as statistic is animate, believing: “We do kind of live by averages and classify students as above, below or on par, and that doesn’t always give me

10 All quotes derive from interview transcripts, field notes, and correspondence with the 10 principals during the December 2012-April 2013 interval. See Appendix H.
answers as to how I can change that, but at least if you are aware of certain things.”

Principals view neither a measure of central tendency, nor even the most detailed representation of student performances, as a complete picture of a student or of their school.

Rather, the averages within a data display serve as cues to principals at specific points within the provincially-mandated cycle of planning and reporting, that governs their managerial practices. As index, the average is a point of departure for reasoning and action. This point of departure is founded on clear premises—that principals’ leadership role is to both sustain and improve student learning while also maintaining a balance in the way that school staff and community members’ interpret student achievement. As a symbol, the average is assigned divergent meanings by various individuals and groups. Principals feel responsible for balancing the many conflicting expectations for students and the school, hopefully bringing conflicting interpretations and expectations into balance. This sense of responsibility leads them to mediate and manage the contradictory meanings of the average through open and closed questioning.

School principals’ sense of the middle is situational, fluctuating, evolving, and dynamic. Principals do not consider the average as a representation fixed in time, nor do they consider the student that it represents as being stationary. They do not define averaging as a textbook problem in inferencing from a fixed total (Zazkis, 2012), since class and school sizes are in constant flux. The middle is an ever-expanding and contracting zone, depending on principals’ specific intentions in the moment and their overall purposes as leaders. Thus, the average is not a crisp midpoint, at least not to the principal who is reading student marks, but rather an elastic middle zone. Craig, like the other principals, tends to define the average as a vague, ever shifting midpoint in school
practices. When asked whether averages sit in the grey area of a distribution of scores he has sketched, Craig answers:

I do and I don’t know in my mind what white is and what black is [pause]. I don’t know which is right and which is wrong. I don’t necessarily say one is and one isn’t. I just know that that middle is [pause] that middle is where I think my world in school is.

Most school principals (6/10) conceive of the statistical average as a centre-of-balance (Hudson, 2012; O’Dell, 2012). They are not so much concerned with its mathematical properties as with its features as clue and cue that something may be awry in the classroom, home or hallway. When they consider the average as a notation in school, school division, or Ministry of Education software records, they often express frustration that bureaucratic rules or professional beliefs have hardened student averages into unreasonable categories. Principals occasionally override an average from one of these sources and do not enter it on a report if they feel it unfairly represents an individual student’s achievement. However, all principals indicated they will sometimes meet with teachers to ensure that teachers’ classroom-generated marks exhibit an appropriate range.

While most school principals held a centre-of-balance conception of the average, four did not. In his on-reserve school, Rick conceives of the average as a flow, where students:

flow along reasonably the same amount of speed, and you know that would probably be what we consider average, you’d have [a] lower end of that average as far as speed and ability and you’d have an upper end as far as ability to do their work, and then you’ll have like 10 percent who you know they’re going to need a little bit extra and a little more help, and then you have 10 percent who I mean they’ve got it mastered before you’ve done giving the assignment.

Similarly, Robert described the average student as “just somebody that just kinda comes and does that business and moves along with, I guess, what the grade-like expectations of the curriculum are.” Phillip thinks of the average as typicality or
representativeness: “To me an average is looking at a big pool of things, it’s looking at a whole bunch of different compartments and now trying to come in and say, what’s typical I guess.” On the other hand, Sr. Margaret conceives of the average as a sign of trouble (Groth, 2005), as signal-amidst-noise—but this noise is in the student’s life, not in the school. For her, averages are often omens or symptoms, signaling the quality of the teaching underway:

The level of academic excellence of the students . . . it can point out that there might be some issues with just where the students are at. If it’s a student that does well in another class, I would be wondering why they’re not performing in this one. I usually look at home situations if a student isn’t performing, try and dig around there a little bit to see what’s happening.

According to Craig, a centre-of-balance conception of the average is a better fit than the fair-share approach because it more accurately represents the many considerations that influence principals’ decision-making. A fair-share conception accentuates struggles over resources: students who are below average often receive more resources. Actually, students who are at average and above average often receive less professional attention, Craig acknowledges. The administrative leadership team in his school division has only begun to consider questions about the strategic allocation of resources in light of outcomes, with the accompanying moral issues, and he suggests that much more discussion is needed before consensus is reached.

Nearly all principals consider the average as a periodic point of comparison and a reference point in terms of specific “compartments” or “strands” of a student’s growth as delineated by curriculum guides or as expressed through varying social expectations. For example, Louise considers the average student to be a composite of all areas of student life; the traits she ascribes to this construct are “age appropriate,” “on par with good social, emotional growth” that is “peer appropriate,” and “academic achievement;
physical well-being; problem-solving and conflict resolution skills,” where the student is “thin” and needs “guidance in some regard.” Phillip defines the average student in terms of “compartments” which together create a more holistic view of the student’s progress through elementary school. Robert too considers the average student as an amalgam of various components.

Therefore, these school principals presumed that multiple meanings for an average develop not because the statistic has multiple elements, but rather because achievement itself as “object” is multidimensional within Saskatchewan curricula. School administrators often defined their role as overseeing the development of the whole student—and each student is an amalgam of curricular and social expectations. For that reason, it would be a mistake to consider school principals’ conceptions of the average as congruent with central tendency, because they did not see individual student achievements as tending toward a midpoint. They recognized that there can be a bi-modal distribution, and that results can skew in different directions. There is no such thing as a typical student, Sr. Margaret insisted, and no centre to any group of students.

Principals did not conceive of the average as an algorithm for computing a mean, mode or median in a multiplicative or procedural sense. Manipulating a particular statistical form is a practice they would not find meaningful, nor for which they would find time. Very little manual calculation of marks is involved in school principals’ work. They authorize student achievement and are not statistical authors. Principals usually receive and consider the averages generated by others—from classroom teachers or the school division office, for example. Therefore, they take the average as a given at a particular instant in time, reasoning from it rather than re-computing it. In other words, principals’ conceptions and reasoning patterns are better construed as a form of organizational rather than statistical problem-solving. Principals’ calculations are of a leadership rather than of a mathematical kind.
School principals’ categorical schemata are revealing, particularly when analyzed in terms of Peirce’s sign elements. When classifying the object, principals most frequently deemed the student(s) as above, at or below average for all three of the data displays they read prior to their interviews. With few exceptions, they did not classify students or ask questions about them according to equity categories. Gender, socioeconomic status, and Aboriginal or ethnic origin were generally not of interest. When principals were considering a data display in terms of achievement, they gave primary consideration to the categories of instructional program and grade level. Language arts and mathematics were the programs of primary interest. Averages for health, arts education, social studies, and even the sciences were not considered to be autonomously worthy of action by most school principals, and only of interest in consort with language arts and mathematics. Attendance was considered to be an important variable. What were not viewed as particularly relevant were community accomplishments, athletic feats, or musical performances that are often considered to be forms of student attainment or achievement. The focus remained on the core curricular program, not on extra-curricular activity.

When considering the average as a Sign or representation, principals’ usually categorized averages according to their source—whether the averages were generated by the classroom teacher, through a computer software system, by a central office consultant, or through a Ministry of Education or other external assessment. Principals did not classify averages by statistical types— the mean, median, mode, and standard deviation as identified by the statistician-author. Principals’ classifications for the average as form revolved around its provenance rather than its mathematical properties.
Categorical schemes for the interpretant remain difficult to characterize, given the impossibility of disassociating a particular data display from the interpretive process itself. Nevertheless, in interviews, school principals tended to classify averages in terms of temporal points in a school’s yearly cycle. At the secondary level, a mid-semester mark invites a different reading than a semester-end average. At the elementary level, a report card grade is considered differently than a grade which is read as part of a monthly action planning process. As Sr. Margaret notes, report card time, unlike other times in the school year is a high-stress time within schools for staff, for students, and for her.

Saskatchewan school principals therefore conceive of at least three types of averages in relation to student achievement. As outlined in Table 6.1, the source is important for types of average: (1) those set by the teacher for individual students’ report card marks; (2) those calculated across a classroom or grade level, which may be generated within or outside the school, usually through software; and (3) those which compare the entire school to other schools, and are calculated outside the school. For the most part, student report card marks are principals’ primary concern; they pay less attention to classroom or grade level averages. However, there is increasing concern for the whole school average in literacy and numeracy as a source of both community pride and pressure which the principal must both represent and mediate. When these various measures of centre do not align, school principals often engage staff and others to moderate and reconcile their divergent implications.
Table 6.1

<table>
<thead>
<tr>
<th></th>
<th>Individual Student</th>
<th>Classroom or Grade Level</th>
<th>Whole School Level</th>
<th>Comments on External Assessments</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Tremaine</td>
<td>Weighted arithmetic mean according to curriculum</td>
<td>Mid-range</td>
<td>Percentile with bar set at 50th percentile</td>
<td></td>
</tr>
<tr>
<td>Kirby Klein</td>
<td>Weighted arithmetic mean according to difficulty of test items and district policy at secondary level</td>
<td>Meaningless arithmetic mean</td>
<td>Meaningless in short- and mid-term</td>
<td>Meaningless in way they are administered in province</td>
</tr>
<tr>
<td></td>
<td>Criterion-referenced 1-4 at elementary</td>
<td>Disconnected for school in district software package</td>
<td>Grade 12 Provincial Examination blended score</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Modal 3 at elementary</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Louise Simpson</td>
<td>Weighted arithmetic mean according to curriculum priorities</td>
<td>Modal distribution</td>
<td>Mid-range with floating bar</td>
<td>Arithmetic mean</td>
</tr>
<tr>
<td>Moira Gullickson</td>
<td>Weighted arithmetic mean according to essential skills which have longevity in value for students</td>
<td>Median</td>
<td>Mid-range</td>
<td>Rescaled stanines with grade level equivalent</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Fountas-Pinell grade level equivalent A-Z (lexile)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The Whole Class Reading Assessment (Grades 3-8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pammy Reid</td>
<td>Criterion-referenced 1-4 with deliberate resistance to consider a percentage average</td>
<td>Meaningless</td>
<td>Meaningless unless criterion comparator is provided</td>
<td>Fountas-Pinell Grade-level equivalent A-Z (lexile)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>The Whole Class Reading Assessment (Grades 3-8)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Peter Getz</td>
<td>Weighted arithmetic mean according to curriculum priorities</td>
<td>Algorithm set by software-arithmetic mean</td>
<td>General distribution of high performing and low performing students</td>
<td>Algorithm set by Maplewood software</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Mid-range</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Phillip Brunswick</td>
<td>Weighted arithmetic mean according to curriculum priorities at high school Criterion rating scale of 1-4 for elementary</td>
<td>Mid-range within core subjects but resists as personal information for student and family</td>
<td>Proportions of students according to percentiles</td>
<td>Algorithm set by district software</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rick Gabriel</td>
<td>Weighted arithmetic mean according to Sask curriculum outcomes and teachers’ individual, principal-approved reporting scales</td>
<td>Modal distribution and three standard deviations</td>
<td>Rate of student flow across school by curriculum outcome rather than grade level</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robert Hanson</td>
<td>Weighted arithmetic mean according to amount of classroom time devoted to each element</td>
<td>Arithmetic mean</td>
<td>Arithmetic mean</td>
<td></td>
</tr>
<tr>
<td>Sr. Margaret Westlake</td>
<td>Weighted arithmetic mean</td>
<td>Range</td>
<td>Gaussian curve</td>
<td>Symmetric Gaussian curve but unreliable scoring</td>
</tr>
</tbody>
</table>

Note: All forms identified in principals’ commentary rather than three data displays provided for interpretation in December 2012- April 2013 interval.
6.2 Index and Symbol

If a general rule may be advanced, it is that school principals’ opinion about the role of averages, as symbolic in educational affairs, does not impede their drawing implications from the statistic as an index for their students and their school. As much as Phillip and Pammy recoil against the traditional and outdated connotations of a percentage grade, and as much as Sr. Margaret resents an average as a label which may be attached to the student, all three were ready to question a data display and then other people about what the data display indicates about the student or school. “. . . to us I think [the average] is a comparative bar,” states Craig. “If our graduation rate in our school was below the provincial average, I would be wondering what we need to do differently and then try to make some changes to do those things differently.” Robert likens the average to an “old-school mercury thermometer” that “goes up and down,” a measurement index that is used for self-reference by the principal.

Principals sometimes acknowledge the importance of the home and community context on student learning, but do not consider these influences as predetermining the average or diminishing their role as school principal. If principals have a criticism of averages, it resides in the notation’s over-simplification, in its reductive properties, or its failure to communicate complete information to an audience without a broad view of context. Principals do not generally see context as determinative, but rather as illuminative—in that way, most principals fit within an Enlightenment orientation to data use (Weiss, 1979), but at the school level. Both Peter and Craig will readily identify particular segments of their school community that have a lower socioeconomic status than others and indicate that these students tend to have lower marks. According to Robert,
There’s a lot of excuses that are built into a student not learning. But at the end of the day, if your kid is [pause] the kid is performing to a developmentally appropriate level, then you’ve done your job [pause]. No matter where they came from, no matter what their gender is or anything, I think there’s just a thousand [pause] so many factors that go into it.

Yet not all agree. Kirby believes that an almost infinite variety of local variables are beyond others’ knowledge and hence the generalizing properties of an average are severely limited. “To use the word average and trying to make comparisons from that again, unless you know the context and understand the context, it’s meaningless,” Kirby declares. “And can also, worse, worse off yet, can cause people to draw some invalid and wrong conclusions.” For Moira, in contrast, the context in which an average is computed defines relative success: “If the context is talking about grades, I would think of the average as just being a benchmark for pinpointing your success or failure by comparison,” Moira states.

These are evolving conceptions for the sign, an important feature of Peirce’s semiotic model. A month later, upon reflection, Moira describes the average as “a representation of a norm,” and much later as a “driver for professional action,” assigning it universally to “the median” by which she means a general middle rather than a precise statistical point in space. Louise’s conception of the representation is dynamic: it is an oscillating horizontal bar. For her, a “below average” statistic is “a call to action,” impelling her to direct individual attention to students whose achievement is below-average, while average student achievement is synonymous with “not an issue”. For Pammy, as shown in Figure 6.1, the average evokes different images in different interviews—as a circles superimposed on a four-bar set of criterion levels, as a benchmark level on a vertical bar, and latterly as a fluctuating normative wave contrasted with the longitudinal growth evidenced in criterion-referenced measurement.
Figure 6.1. Pammy’s depiction of average as imposition on four criterion levels (Interview 1); as normative curve and as benchmark (Interview 2); and as fluctuating longitudinal line in apposition to criterion-referenced measurement results (Interview 3).

Principals are intensely aware that others will interpret the sign (the average) and its many objects (students, school, staff, instruction) in ways that do not align with their own. In other words, while principals may not be aware of Peirce’s theory, they are aware of the parallax in sign relationships implicit in his theory of interpretation (refer to Figure 6.2). Moira’s comments illustrate this point: “When I was a teacher if I put a ‘fair’ on a child’s report card, to me that was a red flag to the parents, like you need to be aware of this. So to me, a ‘fair’ was failing, and to the other parent, a ‘fair’ was OK.” As an elementary principal, Moira equates average with “reading at grade level” as minimal competence. She believes that “if I’ve done what I needed to do as an administrator with instructional leadership and shared leadership and if they continue doing what I’ve laid out for them [teachers] when I return the next fall, all students should be reading at grade level, so that’s what I’m interested in seeing.” Similarly, Pammy resists numeric averages, forecasting they will be eventually superseded by criterion statements not only because she believes rubrics better convey student quality, but also because many students and parents misconstrue percentage averages as “saying that the student is either making it or not, and that some are deemed more capable than others.” Moreover, she believes that the Ministry of Education uses averages for comparative purposes to label some schools as succeeding, and some groups of teachers and administrators as more
capable than others. Pammy has implicitly reassigned the object of the percentage sign from student to organization, but the two are so entwined as to be indistinguishable in her mind. The solution, she believes, is to expunge averages altogether and not attempt to elaborate on their meaning.

Phillip also believes that a percentage average is inappropriate for elementary grades, because students at earlier grade levels do not understand the concept. The average therefore does not provide meaningful feedback to students whereas a four-point scale does. Policy changes that enable shifting expectations for grade level performance make the average problematic for Moira. “Tricky” is the adjective she uses when describing how grade level expectations for literacy and numeracy can be adjusted in policy. The continuum that underlies the range of student achievement can be unilaterally changed by central office officials, as it was in her school division, so that what was deemed adequate one year became inappropriately low the next. For school principals, then, the average’s form is elastic and situational. Distinctions between mean, median and mode are not often drawn, although the individual student report card mark is consistently conceived as a weighted arithmetic mean. That is, the reporting form is not necessarily polymorphous, but its meaning is for the variety of parties who must interpret it.

For Sr. Margaret and Pammy, the average symbolizes a systemic imposition on public education, an administrative obligation laid on schools and educators which inhibits instruction by labelling students and affecting their self-esteem while also depriving educators of valuable time. Yet Sr. Margaret is ambivalent about the statistic. “You know it’s like black and white, you need them and you don’t need them. You have a use for them and on the other side they’re a hindrance. It’s a funny little character,” she concludes, thereby personifying the arithmetic form. She sees average grades as an
obstacle to teaching, but on the other hand, they are “a window” for peering into classrooms. She believes that contemplating the differences between averages can be a source of professional growth for teachers, and she thinks that averages must be considered carefully at report card time because they are a predictor of parental phone calls.

School principals’ simultaneous, conflicting reactions about the average is manifest in many ways: it is a point of conflict. As he illustrates in Figure 6.2, Craig explicitly upholds an expanding and contracting grey area that enables him to negotiate controversy within a staff or with an individual staff member over a student’s standing.

Note: Arrow = range which he defines as average student achievement.

Figure 6.2. Craig demonstrating elastic grey area that he mediates, with his diagram on desk indicating shift right in public expectations.

For Louise, the ambivalence in public expectations about whether average signifies “not good enough” or “unacceptable” performance means that educators have the responsibility for defining the midpoint. However, she believes it is the principal’s
obligation to interpret in an ongoing way the location of the midpoint, sorting students above and below it in her Watch List. For Phillip, the difficulty is negotiating with staff when allocating and distributing resources among various groups of students and teachers. The pupil-teacher ratio has more immediate impact on his decisions than does the notion of average student achievement. For Pammy, the complexities of disentangling values and expectations from data, and the misrepresentation of values in an average interfere with the school’s educational mission to demonstrate progress according to curriculum outcomes. To her, it is far better to articulate the curriculum values that are sought in criterion statements, than to let others use numbers for comparative purposes. Rick notes the “toxic” relationships that can arise in sensitive discussions among teachers who assert their autonomy and purview over grading practices, but will attribute problems in student learning in their own classroom to other teachers at earlier grade levels. Sr. Margaret insists that time pressures necessitate a quick scan of marks for anomalies which may become educational or administrative problems rather than focusing on the centre per se. In general, principals recognize that pressures exerted by the Ministry of Education and central office will elevate averages. Most principals are convinced that providing individual administrative and instructional attention to students whose performance is below the situated average will elevate the school’s overall average and will ultimately be beneficial.

To facilitate interpretation for themselves and for others, school principals often imbue numbers with colour. Statistical averages are often conceived as black and white (Louise), as red flags (Robert), as red/yellow/green overlays in the documentation that Pammy creates and shares with staff, and as a grey area (Craig) where there is latitude or leeway. As Craig notes:
I really struggle [pause] I really struggle with completely black or white on issues because I see so much grey in the course of my day, and try to convince teachers they need to see yellow in some cases with average marks, where I see shades of grey because a kid coming into your class might act one way, and if they are coming into my class might act a completely different way, depending on even the relationship with them.

Describing averages in terms of colour is a habit that principals have for managing the meaning of averages. For Craig, yellow conveys a cautionary note to staff. Colour substitutes for the qualities that are perceived as missing from the black and white statistic as icon: adding a colour thereby transmutes the numeric sign from a quantity to an organizational quality. Each colour has its own meaning for school principals, which they superimpose on the notation’s icon(s) to render it intelligible as an index, for direction. Averages are frequently considered to represent a zone wherein controversy may arise. Some principals, who would prefer to see this zone as a grey area in order to accommodate differences within the various “strands” of an average student as object, invest colours in their interpretation of an average, as a way of facilitating meaning. While colour may have its own grammar as a semiotic resource (Kress & Van Leeuwen, 2002), this theme in school principals’ commentary is consonant with Peirce’s propositions about abductively imbuing a numeric icon with metaphoric qualities of resemblance, to create a possible rule (EP2. 395-396).

Most importantly, school principals’ conceptions of average student achievement are sometimes articulated in opposition to the idea that society considers an average as foreordaining membership in a particular social class. As has been long recognized, nearly all principals in this study acknowledge that student grades as averages can be viewed either as tools or tyrants (Briggs, 1964). Although the association between two different dimensions for the meaning of a sign—the average as summative measurement index and as a socially-potent symbol—must necessarily remain impressionistic, their
commingling fuses two different elements in Peirce’s semiotic. Notations of average student achievement clearly serve categorization functions in principals’ minds. Given Craig’s comments on pride and pressure, Sr. Margaret’s concern with long-term labelling, and Pammy’s concern about miscommunication in reporting, we might further consider the interpersonal and hence organizational idea that the average often functions as a “status symbol.”

6.3 Gestures and Embodied Ideas

One of the most telling expressions of school principals’ conceptions of the average is their often unconscious gestures while speaking. Whether it is Kirby’s verbal description of how student averages are created by merging first-term and second-term assignments, which is accompanied by bringing his hands together to show the interlocking scores with his fingers to form the average, or Sr. Margaret’s tracing the outline of a normal distribution in the air and then making a cutting line through the centre in a sort of obverse sign of the cross, school principals’ ideas about the statistic are usually accompanied with motions of arms and the pointing of fingers. As shown in Figure 6.3, Louise positions her right arm perpendicular to her body to illustrate the average, using her left hand to make a sorting motion above and below, pulling the below average toward her. Robert initially sees the average “as a line and we’ve got to get across it. We’ve got to get over it somehow,” tracing it horizontally in the air while verbally likening it to an Olympic minimal qualifying requirement. Gesturing to show the midrange with upwardly cupped hands, Moira conceives of the average as a “pocket” into which most students fall, whereas Phillip cups his hands in the opposite direction to illustrate the peak of a pile. For Craig, the distribution of student marks is typically skewed right with the majority of marks above a 65 percent point, but this average remains mobile within that distribution. When explaining this conception, he expands
and contracts his hands to illustrate the flexibility of his mid-range. What is
unconsciously expressed as index becomes a sign of quality for me—my first impression
is of the quality of Craig’ statistical understanding, with a second realization that his
gestures also serve as an index for giving direction to staff and students. Thus, Peirce’s
claims for the multivalency and anthropomorphism of mathematical ideas are readily
apparent in interviews (CP2. 322-324), illustrating the way mathematical forms become
local norms.

![Image of a person gesturing](image.png)

*Note:* Arrow = direction of arm movement.

*Figure 6.3.* Louise triaging students for her Watch List with right arm as average bar,
and left hand scooping below average students toward her.

When the statistical idea is expressed as a gesture, the sign is transformed from a
written notation or electronic image to bodily form. What from my point of view is a
gesture and, as such, an icon of a statistic’s qualities, may not be perceived by the
principal in this way, but rather as symbolizing a statistical concept with inherent rules.

Gestures may simply be another way of articulating the average’s meaning concurrent
with verbal expression. For Peter, the average sits as the centre point in his person, his
hands indicating above and below average, his fingers pointed either up or down to signify individual students and their grades, and the scale balancing in a vertical line down the centre of his body. His finger represents his indexical authority within the school, and for me, is a scale of educational justice over which he presides. Peter admits, upon reflection, that he had never considered the possibility that in his person, he represents averages as a school norm. In a follow-up email, he reflected:

After you left, I thought more about our conversation, particularly the part where I used my hands to illustrate an average. I have never thought of an average as a balance but rather a mean. My hands on either side of my body would illustrate the extremes of marks—high and low. My body would be the average or mean. I’m not sure how our conversation evolved to the balance idea. Perhaps I should have turned my palms in instead of up. Hope this makes sense.

Principals are often unaware that they have expressed statistical or normative ideas through their gestures simultaneously with their words. They are not conscious of this action, drawing into question one of Peirce’s phenomenological precepts which emphasizes capturing what is in the consciousness. This suggests that statistical ideas may be more primal than might be suggested through psychological schemata or theories that revolve around encoded images and information-processing which presume the mind operates like a computer hard drive. Using gestures to explain normative ideas may be viewed as creating a visual (re)presentation of internal mental life (Streeck, 2009) for an audience. From my perspective as interviewer, gestures are an iconic depiction of thought, or an articulation of a psychological concept that arises from the periphery of attention. But from an embodied cognition perspective, the conception is initially a bodily expression of a statistical idea which thereafter becomes a feature of mind. At

11 Such gestures may be neither performed for an audience nor simply the consort of verbalization. I recall in 2004 observing a distinguished North American measurement specialist, Dr. Ron Hambleton, unconsciously mime an entire technical report, while reviewing the complex calculations within item response theory, as he was surreptitiously reading it at a table corner during a Toronto colloquium while other speakers held the floor. Hambleton (2002) has spoken and published widely on issues relating to the interpretability and use of student achievement data.
issue is whether reading elicits more primordial, bodily notions of number, where the manual precedes the mental and the physiological precedes the psychological. Does the thought come first and generate the gesture, or does the gesture come first and cause the thought? Regardless of how we answer that question, most school principals, through their gestural expressions of statistical ideas, articulate an average as a centre of balance, often suggesting in their posture and action that average student achievement may be expressed kinesthetically. When such gestures are consciously performed, they may be defined as an element of charismatic leadership, but when made unconsciously, they may be considered as part of customary or traditional leadership (Holladay & Coombs, 1994; Küpers, 2013). Gesticulation is now considered to be the leading edge of cognitive development (Roth, 2000, 2001) and was integral to Peirce’s anti-Cartesian stance (EP1. 226).

In light of the foregoing, we may therefore accept the conjecture (1a.) that school principals’ centre-of-balance conceptions of average student achievement are nuanced and multiple. Their conceptions are expressed in verbal, diagrammatic, chromatic, and kinesthetic forms that are distinct from various mathematical and statistical formulae as defined by a statistical author. Principals may criticise the average when the sign does not conform to their own beliefs about its object (students) and their convictions about appropriate ways to measure achievement. Principals personify a local norm, and their gestures often represent their ideas about the average.

6.4 Sociological Norms and Mathematical Properties

School principals recognize that they perform normative roles in their school and in their community. However, their notions of the average are situated in disparate views about normality, where various forms of the average follow and fit with administrative function. For most principals, that norm is local, although most are quick to say if their students did not meet a national or international average, they would take action.
Kirby’s outlook, the average “gives people something concrete to hang their hat on. It takes the abstract and puts it down to a very concrete, simple—maybe not simple, but recognizable form.” With his background in the physical sciences, Kirby readily distinguishes between a mean in the natural sciences and one in social sciences, which he believes necessitate different conceptions of the norm. Sociological norms are abstract attitudes and masses of human behaviours made concrete, Kirby argues, whereas physical science norms are concrete things made theoretical as mathematical abstractions. In education, the student object changes over time, whereas in engineering, the material object is often stationary in its properties. Kirby states, “whatever the majority of the people believe concerning that topic becomes the norm . . . that is why from a sociology [sic] point of view, this whole area of behaviour and attitudes—the norm changes as the decades go . . . because attitudes change.” Kirby thus affirms but does not draw on Peirce’s dictum that norms are best observed over the long run, holding a frequentist view of probability similar to Peirce. Sr. Margaret is clear that her conception of the average is not related at all to faith: the systematization of education and the sorting requirements of post-secondary institutions and not encroaching secularization explain the use of averages in her school. For her, the average as applied to students reflects a systemic constraint and propensity to label which stands in the way of a more direct exercise of students’ will to learn. She is certain that teachers can create that will to learn within students through the exercise of their professional will, and by knowing where students are in a continuum of learning.

School principals’ conception of the average and of variation thus underlies their sense of organization. Yet it would be a mistake to characterize principals’ approaches to establishing norms as examples of Peirce’s *a priori* method, where the fixation of belief is overly rigid. Several different perspectives on variation have thus far been identified in experts’ statistical reasoning (Peters, 2009)—a design perspective which emphasizes
expected but explainable and controllable variation; a data-centric perspective which revolves around unexpected noise around a signal; and a modelling perspective where variation is expected and deviation from expectation is ascertained. In this study, nearly every principal interviewed exhibited a design perspective to variation in statistical reasoning. The issues they raise in sample representativeness (for both classroom evaluation and large-scale assessment), in the weightings assigned to classroom assessment marks, in their concerns and direct interventions to ensure that teachers’ marks display an acceptable range, were all explainable and susceptible to principals’ influence.

These classroom grading issues in representing variation involve the “nimble numerator” from shifting weights, or “deficient denominators” by not including sufficient measures, or both, and have many concrete implications for school administrators. Kirby explains that a “teacher can literally create any average mark that they [sic] wish through shifting weights and also through the [test and assignment] questions you ask.” He considers it important that a “documentary trail” be left behind so that teachers may defend their assessment procedures and their judgements, or so the principal can. He describes recently retrieving a teacher’s marks book over the summer holiday to address a complaint. After successfully defending the teacher’s grade, Kirby subsequently talked with the teacher about “an issue” that Kirby saw in the marks book.

Likewise, Peter will occasionally convene a one-on-one meeting with a teacher in his office when the marks assigned by the teacher are unacceptably high or low, having glanced at marks assigned by other teachers in a variety of subject areas. During our interview, he discussed the subjectivity of marking, describing how he would ask a teacher to “look me in the eye and say that you know absolutely with certainty that that student deserves this mark,” and indicated the “human error” that can be involved in
student evaluation. With these comments, he seeks to create doubt while also revealing his own. “So the point I’m trying to make to the teacher is, ‘Help me!’” Peter concludes these infrequent conversations with such advice as “I need you to think of some things that you can do differently in the next term; whether it’s give more tests, requiring the kids to remember less stuff . . . so we can bring these marks up to where they need to be, or conversely if they are really high.” In a recent case, “a constant conversation” ensued after the crucial conversation, where the teacher approached Peter frequently about his evaluation practices. Sr. Margaret sometimes teams an aberrant marker with another teacher to rescore student work so that marks are “moderated,” as does Rick. Kirby’s preferred strategy is to coach neophyte teachers in the art of constructing tests with a range of question difficulties that will yield a defensible distribution of marks.

Therefore, the marks become normative when the principal intervenes to ensure the range falls within the bounds of what the principal would deem as acceptable to outside audiences, when aligning the marks in comparison with attendance data, or when judging whether a teacher’s marks conform with the principal’s understanding of an individual student’s ability by looking across the span of a student’s report card marks. This suggests that students may sometimes be graded in terms of their perceived ability, not their actual performance, and that a student who is labelled as being of low ability will be subsequently graded above or below average because of that label, regardless of their actual performance.

Some principals question whether a data display is factual at all. According to Rick, for example, averages are a conjuring act “to make real” a largely invisible process. When comparing his teaching and administrative experience to his previous experience
of marks when he was a student, Rick now understands that assessment is “like when you see the magic act from behind the curtain, you get to see the tricks in it.” Pammy is knowledgeable about John Hattie’s (2008) Visible Learning, a transnational meta-analysis of factors which will improve student achievement, but believes that it is criterion assessment and not a percentage average which makes learning visible and actionable. Similarly, Kirby is adamant that averages are meaningful only within the community that generated them, and that unless you understand how they were produced, the decision maker will go astray:

And that’s what I mean by the dangerous side of this. If you don’t truly understand from where that number came then the potential is certainly there, because these numbers are your decision making [pause] lots of decisions are based on the numbers. But if you don’t truly understand from where the numbers come from and you just give it a blanket number and you might have this impression of what this number means and you’re going to make a decision based on that number and what your impression is: well you may or may not be right, and your decision may or may not be the correct one.

His notion of a fact is as a visually-supported belief shared by the community, a fallible computational demonstration introduced in a persuasive process. School principals often do not feel they are acting upon the solid bedrock of fact: the average is rather a provisional position which must then find confirmation, or else he or she will shift in their footing. In a Peircean metaphor, principals are “walking on a bog, and can only say that this ground seems to hold for the present. Here I will stay till it begins to give way” (EP2. 55). School principals have constituencies to mediate, but they do not generally consider a school average as the only tool for mediating successfully with those constituencies. To do so would be to acknowledge that the statistical average holds power over students’ and teachers’ destinies, and would explicitly politicize the principals’ role.
Therefore, most principals see neither classroom grading nor large-scale testing as a state-sponsored sorting device. Although Rick, in his reserve school, sees numbers as having a bureaucratic function for classifying and categorizing, he does not equate them with the machinations of the Department of Indian Affairs (now known as Aboriginal Affairs and Northern Development Canada), nor the designs of either provincial or federal governments. Rather, Rick’s antipathy toward a percentage average as statistical representation resides in its failure to provide sufficient information for focused action. More importantly, attempting to have all students performing at an average level or above undermines students’ sense of belonging. His concern is that a below average label in some subject areas undermines students’ sense of efficacy and hence their engagement in his on-reserve school:

What we really struggle with in our school is to have students feel a sense of belonging no matter what their status is. We want them to think this is their school. We want them to feel good about coming here. We want them to feel that there’s a positive thing coming here.

Peter admits that when it comes to student averages, the “buck stops at his desk,” and that he takes the responsibility of monitoring student marks seriously. Phillip’s approach, on the other hand, is to “go along to get along” when it comes to controversial issues such as formulating a student evaluation policy with a superintendent, or officials further up the administrative hierarchy. His impulse with people who “rock the boat,” or with those who want every issue addressed, is “This is what we have to do instead of fighting it.” Phillip uses words, gestures and a diagram, as shown in Figure 6.4, to express his ideas about the norm, his normative influence, and his avoidance of instructional panaceas. Most importantly, he sees himself as someone who represents the norm—the middle—as someone who is adverse to extremes.

D: Do you see yourself as representing the norm?

P: The norm? I think in a lot of ways, yes. I think in talking to other administrators or talking to other teachers, I do both. You know I teach part of the time and am an administrator. I don’t see myself as doing anything
that’s way out there. I don’t see myself as being a person that’s got these drastic measures that you’re completely trying the same thing, so yeah I would almost say yeah, I’m kind of probably in the norm, kind of along the middle, the average.

![Diagram of average student achievement.]

*Note:* A = average student achievement.

*Figure 6.4.* Phillip’s diagram of average student achievement as a neutral middle between good and bad performance.

Konold and Pollatsek (2004) point out that computations of average and conceptions of variability are inseparable. Reasoning about averages and variation merge with reasoning about the spread of data relative to centre, as a form of distributional reasoning: the two together are indicators of sophisticated statistical thinking (delMas & Liu, 2005; Groth, 2005). Among all principals, if there is an underlying image of a statistical norm, it is as an iniquitous bell curve or gradient that is often described as “old,” “traditional,” and “conventional.” For Sr. Margaret, like Craig, the average is the centre point of a Gaussian or symmetric normative distribution. In Craig’s analysis, our older notions of a norm have shifted dramatically with technological change and the need for knowledge workers, rather than manual labourers.

It used to be that you did have your general group of students in the middle of the bell curve that I’m drawing here . . . but you did have the smaller percentage of
kids at the top end that were shown on that as successful and the low end it was just accepted for the longest time that that low end was OK. That there was going to always be that low end of kids that were going to be below 50, but in general you were going to have a bell curve settle over the 50 percent model. They’re saying now that in this part of the world or this time of the world that’s not accurate anymore. That anybody below this 50 percent percentile mark are people that aren’t getting the jobs anymore. There’s no jobs for them, nobody in the technical world, . . . there’s not enough labour intensive kind of jobs for the less academic but this is what the new curve is supposed to be.

He thus recognizes that rising classroom marks over time reflect but do not necessarily represent those socioeconomic trends. In fact, he readily conflates a report card average, as summative index representing a student’s standing, with the range or mean or mode within a distribution of classroom or school marks. For Robert, in contrast, the average as represented in a student mark is different from a norm. Robert sees norms as invested with social acceptability—an imprecisely defined public’s view of what is acceptable, whereas averages are more objective values that are simply manipulated in the number, to find the middle. Kirby considers 75 percent as about average at the high school level, and a 4 on a 5-point scale as about average at the elementary level. A normative distribution is thus a fairly accurate reflection of student performance in his school, Kirby admits, but with a small school, it is not worthy of either his administrative time or teachers’ professional time to formally plan toward improved results.

Principals’ notions of a statistical norm remain elastic, in part, because of their own teaching experience in constructing a range of marks, through weighting in the denominator. They receive marks from central office, from teachers, or in Peter’s case from a district software package, as a given and usually do not re-calculate them. For the most part, the averages provided by teachers for individual students are described as the weighted arithmetic mean, which is constructed out of students’ assignments, tests and other elements throughout the school term. Such constructions are considered
idiosyncratic to the individual teacher, although some schools are trying to bring consistency, if not uniformity, to the elements that make up a student’s average. As Phillip notes:

Let’s say it’s a high school student, and you’re figuring out averages for English Language Arts class in Grade 11. The value that you place on different assignments will affect the averages. The value that you place on testing is going to affect your average, so let’s say one teacher puts 20 percent value on all tests, for their course. Now let’s say another teacher who teaches the same grade, the same class, puts 30 percent on the value of tests. Those teachers are not going to come up with the same average, and yet they’re teaching the same course to the same age children.

The solution for Phillip is to work toward having equal weights put on the various kinds of assignments, a consistency that he reaches by working collaboratively with teachers with data.

Principals’ notions of the statistical norm often remain vague, but they have clear ideas about normal behaviour. Attendance and those affective qualities indexed therein are not incidental or peripheral to their conceptions of an average student, but rather are infused within a more holistic view of achievement. If there is a persistent lament in school principals’ commentary on average student achievement, it is what they perceive to be the folly in school board and provincial policy of ignoring students’ behaviour and other attitudinal elements when computing students’ grades. They argue that the trend toward removing these components from evaluations of academic performance has contributed to the gradual inflation of student marks, created behavioural issues that require principals’ attention, undermined teachers’ professional prerogative, and diminished expectations that students attend school regularly. In Kirby’s view, such school board policies have, in effect, removed an important element of classroom control. “There’s a set of normal social skills that people should have when interacting in a public place,” Kirby explains, “and jumping, skipping, flipping and flopping all over the floor is not one of them.”
Not all school principals subscribe to claims about changing social or educational norms. For Pammy, the average is associated with norm-referenced measurement in the United States, with American minimal competency testing, and with annual reporting requirements. In her estimation, all three impede demonstrations of curriculum outcomes. To her way of thinking, averages stand in the way of gathering a wider variety of evidence to demonstrate student learning. Many school principals express reservations about the average expressed as a percentage, seeing it as a traditional means for calculating a student’s standing in a class. It is not the scale itself that is the obstacle, but rather its austerity in not conveying sufficient information. For Pammy and for Sr. Margaret, a percentage average just does not convey where students are in their development, nor what they can actually do in relation to curriculum outcomes. As Sr. Margaret states:

SM: I’m not sure, I’m honestly not sure, I just know I’m not sold on marks because I see kids change and I see kids come alive. I think sometimes averages and report cards are more to their detriment than to their help.

D: How so?

SM: Because it’s that whole thing, they get labeled. You know they have to go home with a report card that doesn’t have that high mark on, thank God! They don’t any more, although we need to put them yet. We have to put an achievement of some kind on there.

Like Sr. Margaret and Phillip, Rick finds “simple averages” to be meaningless without specific information about the nature of the task, about the grade level, or the subject area. “That’s one of Barry Bennett’s [as workshop leader in professional inservice] favourite stories to tell—a kid that was complaining that he got 80 percent in his Math . . . on an essay. ‘What’s wrong with an 80 percent, that’s a good mark?’ he said. ‘Yeah, but it doesn’t tell me why? What do I need to do to get a 90? It doesn’t tell me why I didn’t get a 70.’ There’s no useful information.”
As well as thinking that averages are barren without accompanying information, some principals also believe that percentage averages are fluctuating indicators over time. Improvement can only be identified over a three- to five-year interval, in Sr. Margaret’s estimation, not during the short time frame of one year. She sees the percentage average as a constriction, as an overlay which parents and policy place on top of four basic, criterion, performance levels represented in a scoring rubric. For Pammy, the average is synonymous with a percentage mark, which disguises the range of evidence that may actually demonstrate whether students have met a particular curriculum criterion. It is the percentage scale and not the average *per se* that is distorting and a throwback to older ways, not only of reporting but also of conceiving student learning.

From a mathematical perspective, many school principals’ responses to the average become intelligible through distinctions among the nominal (using numbers to name or identify), cardinal (using numbers to indicate quantity) and ordinal (using numbers to order or rank in a set) features of number (Weise, 2003a, 2003b). For the most part, the principals in this study accepted cardinality, but held different positions about the ordinal and nominal properties. Pammy can reconcile assessment with instruction by accepting cardinality, as the most efficient and least “damage-causing” way of coding students according to criterion principles. What she and many administrators reject is ordinality—using numbers to rank order students. Sr. Margaret and Pammy recoil against using averages in this way, one seeing teachers’ grading as a by-product of the systematization of pedagogy, and the other as imposed by post-secondary education quotas on high schools. Sr. Margaret rejects the nominal features because the number may somehow become attached to the student, through its messages about self-worth, or through the durability of the report card transcript, which will constrain future opportunities. In contrast, Peter accepts the ordinal use of numbers for
ranking students, and is comfortable with his district’s policy of placing class averages on report cards because that provides additional information to parents, so they can see where their child stands in relation to others. In contrast, Kirby has had that feature of his district software package removed when report card marks are generated. Both Craig and Peter believe the rank ordering of students is a necessary part of many secondary school functions. Moira and Robert recognize that ranking is anticipated within any context where student grades are used.

Rick has no difficulty with the nominal or cardinal features of an average in his on-reserve school. He recognizes that using numbers to name or identify are a mathematization of any school population that must stretch a number across disparate groups of students. Averages are part of the bureaucratic function in any school system, and cardinality is key to funding formulae, whether in public schools or in schools controlled by First Nations. The number of students in a classroom, school, or school division figures prominently in most funding formulae. Teachers and school administrators are intensely aware of pupil-teacher ratios and salary scales, and school size is a live issue across Saskatchewan—all draw on cardinal features. We might also point out that although notions of cardinality are intrinsic to the Western tradition in measurement, and may be in traditional First Nations counting practices, the nominal are not necessarily intrinsic to assessment as a process for collecting information. Peirce himself recognized the trivalent properties of a single number when applied to any phenomena (EP1. 245-256), in his mathematics and in his logic (Houser, Roberts, & Van Evra, 1997).

Any conjecture (1b.) about school administrators’ considering the average as a weighted mean in relation to curriculum priorities is therefore simplistic and perhaps misleading. When conceiving of an average, principals bring images of Gaussian curves,
a strong sense of the normative expectations within the school and wider community, and an emphasis on student performance in language arts and mathematics rather than student performance across the entire curriculum.

6.5 **Attention and Simplification**

What causes principals to pay attention to a particular student’s average? Robert describes it as intuitive feeling, based on prior knowledge of how a group of students, often individual students, have heretofore performed.

Something would . . . [pause] my spidey senses would tingle up if I saw a kid getting an “A” and I knew . . . [pause] cause I know what my reading data looks like. I know what my kids are getting, like I said, we have a small group of kids in the school. So I see a kid get an ’A’ in ELA, then I’m asking questions if they’re three or four grade levels below it in reading. I’m asking questions.

In Robert’s experience, it is the discrepancy between the averages among assessments, such as between a teacher’s report card mark and a division-wide assessment average, not the distance between averages within an assessment which he describes as “a red flag.” He looks for those kids who “fall between the cracks, those who don’t show growth or that are below grade level and are not in any individualized program.” When Louise is asked whether she would phone the parents of average students, her reply is: “Well, probably not. (laughter) You know why? Because they don’t cross my desk so much. I guess contacting parents of the average falls into a teacher’s realm.”

The concern that prompts a principal’s inquiry into a particular student’s below average standing is often affect-laden, sometimes stemming from a sense of injustice derived from their experience as a student or teacher. Robert recalls the harangue he received from his own parents when his school marks were not high enough. Kirby conveys a sense of anger and frustration when he recalls the “gong-show” in a university’s attempts to move from a percentage to an 8-point scale for post-secondary
students, robbing him of an 85 score in a course. Sr. Margaret still recalls the satisfaction of receiving a high percentage mark from a sexist university professor. As with Louise and Sr. Margaret, Robert’s first impulse is one of compassion, to give individual administrative attention to “those who appear to be meeting expectations on the outside, but then when you do something with them, it’s like “Holy Smokes!—they’re down, like I’ve got to help this kid.” Yet that feeling is often suffused with a sense of external expectation. Robert as a novice principal believes that his administrative career hinges on elevating student marks, feeling that enormous expectations have been placed on him as a first-year principal. At the same time, he is fearful that because of his small school enrollments, year-over-year fluctuations in student performance will be large. Despite this, he believes that he has been chosen to restore morale among his school staff, to start sharing data with staff in a way that his predecessor did not, and to make improvements in literacy and mathematics, while also mediating the disparate expectations.

If a feeling of dissatisfaction often animates principals’ attention to the below-average notation, it also shapes their reductive approach to reading a data display. To read statistics quickly, most principals focus their attention by imposing three broad and shifting categories—students who are above-, below and at average. This sorting might be considered as type of triage (Booher-Jennings, 2005, 2006), as a cognitive process for selectively concentrating on one aspect of the data display while ignoring other aspects. Whereas the statistical author’s work is often initiated with a narrowly-defined question and culminates in a deductive conclusion within the bounds of the inquiry, school principals as readers of statistics often start with questions and pursue explanatory hunches or diagnoses relating to the classroom or community context. Both statistical authors and readers see contexts as contingent, but the school administrators’ context is more relational than correlational, circumstantial rather than controlled, outside the data display and not within it.
When school principals read data displays, my frequent observation was that the doubt was expressed as a sense of discomfort, but not disconcertingly so. For example, when Louise read Data Display 2 and I pointed out that it could be alternately framed as 60 percent of students meeting or exceeding the average, she stated that although that framing made “her feel better,” she still believed that if 40 percent of students were below average, she would feel that she must take action. If the average represents some point of statistical centre, those statistics which do not conform to principals’ sense of centre provide them with a reason to doubt. This sense is not a precise threshold of (un)acceptability, but rather is located somewhere on a situationally-defined continuum of feeling where action becomes desirable. School principals did not see the statistical average per se as a point of moral conviction, and they were resolute that the average as typical is not an assertion of a probabilistic certainty—none of the principals interviewed could interpret a p-value. However, a below average notation would occasion sufficient doubt to impel them to action.

6.6 Interpretive Processes and Inferences

If conceptions involve cognitive and emotive interpretations for a Sign, then inferencing refers to those patterns of thought that arise with a Sign during the interpretive process. Although principals’ approaches to interpreting a data display are idiosyncratic, I observed several commonalities in the way they accomplish a reading. First, all principals report that they read data in the office while seated behind their desk, from paper or off a computer monitor, and rarely at home. Reading student achievement data is neither leisure time activity nor homework, perhaps because student marks are often considered sensitive or confidential. Principals most often sign report cards one at a time, from a stack on their desk, after scanning each report card for discrepancies, setting aside those that provoke questions. Other statistical reports too are scanned with a similar concern for “things that stand out.”
Second, principals usually read an individual data display quickly, within 30 seconds to one minute. The majority do not read with pen or pencil in hand to annotate, nor with a notepad or calculator, which would slow down the process of interpretation, although more than one principal puts student names on a Watch List or grid for further action. In other words, the inferencing is rapid and governed by quick impressions rather than laborious calculation. Third, the specific procedures or tactics for reading appear to be determined by the display’s format. Most principals started in the upper left hand corner of Data Display 1 and at least half used their finger as a cursor, moving horizontally line by line down the Display. The other half worked both horizontally and vertically, comparing notations in columns or rows, paying attention to column and row headings, making comparisons and deliberately ignoring footnotes and sometimes prose commentary as irrelevant. When graphics such as histograms are provided, as in Data Display 2, principals frequently looked at the gaps between bars before reading axes labels, although some would start with the accompanying prose commentary so as to establish a context. Most ignored footnote definitions for notations such as probability values, although secondary school principals would search for the types of courses that were included in a category. More than half of the principals interviewed sought the total number of students or schools involved in Data Display 2 so as to consider the denominator. In short, principals’ initial lines of inference were suggested by the layout and scope of the data provided. Fourth, and most importantly, the principals search for anomalies, particularly those cases or situations where below average achievement is evident. What they looked for was deviations from the average as notation, not in the statistical sense of a standard deviation, but from their own sense of what is normal based on the practical knowledge they had acquired through experience. This is a form of transactional reading—they did not draw on a precise or encyclopedic statistical lexicon
for interpretation. The inferences drawn were allied with notions of uncertainty and variability that depart from mathematical and probabilistic conventions (Makar & Rubin, 2009) of generalization from or to populations.

What principals seemed to see and seek in the data display were vagaries in the achievement of students in their school and thereby potential shortcomings in their leadership. When they could not discern either in a graphic or numeric notation, they would not find it useful. Revealingly, Sr. Margaret found the gender differences in Data Display 3 meaningful because of her experiences teaching female students, and her previous experience in graduate school, but most other principals did not find this information meaningful. Data aggregated at the provincial level did not become more meaningful just because it carried a Ministry of Education logo—depicting particular students or courses or grade levels did become more meaningful. Disaggregation of data rather than displays of the author’s authority is thus key to the practical significance of the statistic. Statistics were a window into their “object’s” (the student or classroom or school or leader) status in terms of achievement. However, with classroom and school-wide averages, principals did expect a distribution in students’ marks and, based on their knowledge of individual students and teachers’ instructional style, would offer hunches or diagnoses that may pertain to an individual teacher’s judgements if that teacher’s marks were at variance with the marks of other teachers. Principals tend to read statistics as one might read a set of gauges, looking for signs of system malfunction or student problems that require further attention.

In brief, when principals reviewed a sheet of marks, it was in terms of the impact on individual students and on the principal’s administrative practices rather than in terms of the author’s ideas about statistical significance. They looked for general patterns across a variety of subject areas, but particularly in mathematics and language arts.
courses. To obtain a well-rounded picture, principals drew together teacher marks, external assessments, student attendance statistics, and their knowledge of individual students’ circumstances. It was the gaps and patterns among the scores on multiple assessments, rather than the measurement of distance between a fixed centre and outlying scores that had significance for school principals. Principals compared averages across assessments more than they compared means within any particular assessment. The average was not a fixed anchor point so much as a temporary baseline for ascertaining progress in relation to a multiplicity of curriculum specifications. Principals with secondary experience understood that individual student marks have both medium-term and long-term consequences, whereas some elementary principals did not, seeing student marks as more transitorily indicative of immediate issues in students’ development.

Phillip’s commentary on one element in Data Display 1 illustrates many aspects of reasoning with an average:

You see I think with Judith’s [marks] you have to look at a whole lot more than just academics there. I think number one, if you look at the attendance, are they at school? I guess the first thought that would come to mind with Judith is, “Is that an accurate reflection of what Judith can do?” And I would probably say chances are Judith is missing a number of assignments or she’s missing here … so we don’t know what Judith can do. Judith may be a very, very strong student but unfortunately in schools part of your attaining your mark or your standard is there has to be a quality of work that is shown or done so you know with Judith in most cases I wouldn’t even put a mark. I would talk to those parents and say, “You know I don’t know exactly what Judith can do because Judith isn’t here or Judith isn’t handing in the assignments. Unfortunately schools require a mark, OK Judith is a 1 because she hasn’t done anything but is that an accurate reflection of what Judith can do? I don’t think so.” So that would be an interesting conversation to have with that family.

Here, the principal seeks a more holistic perspective on the individual student by collecting evidence beyond just academic performance. When key pieces of information are missing, a question or two arises in the principal’s mind. Then the principal poses a hunch relating to missing evidence and attendance. Thereafter the principal establishes a tentative criterion for the requirements that evidence must meet in order to reflect a
quality standard, with Phillip casting further doubt on the validity of the mark. This culminates in a forecast action that leads the principal to the parent for answers, in this case, with suspended judgement. Intrinsic to this principal’s interpretation is a critique of marks as inappropriate in some cases, but expressed as an institutional or systemic imperative that the principal is reluctant to uphold without gathering further information from parents. This is the logic of abduction, which is difficult to disentangle from making deductive claims, or the inductive search for evidence.

6.7 Credibility and Trust

School principals rarely use the psychometric ideas “reliability” and “validity,” but they often use the leadership terms of “credibility,” “legitimacy,” and “trust” when talking about averages and student outcomes. Credibility is a synonym for believability—that one’s interpretation of events aligns with others’ interpretations; legitimacy is the respect that others accord to the principals’ beliefs. Rick describes legitimacy as key to the management of meaning for an average, because his interpretation must be perceived as right in the community:

If I stand up in front of the community and I present myself in a way that they all reject, all refuse to accept I won’t be doing much good here very long, I’ll lose trust and if they don’t get rid of me, I should leave. [pause] Because that trust is key. Key within the staff, it’s key within the students, it’s very key within the parents. You think about it. The last thing most parents tell their little kids when they start to leave for school is, “Listen to your teacher, listen to your teacher.” They trust us. They may not know why but they’re telling those kids trust them . . . . [pause] Economics; yeah, we control all the things they want and a little bit more gym time, time on the computers, but that’s losing its lustre, that’s not like it used to be. The only real power we’ve got is authority and that comes from trust, and if there’s no trust there’s no power.

When Rick is challenged by parents in public meetings, as he sometimes is for attendance and available instructional hours, he will present every parent with records for attendance, instructional hours, and achievement. However, there is “always a group of people who are saying, “OK, yeah sure that’s the number you’re showing me but is it
legit, you’re comparing this to [a neighbouring] public school, showing us how you guys did.” For Rick, a legitimate reference point involves “a fair comparison, apples to apples.” When asked about appropriate reference points for comparison, his response is:

R: That’s what makes it all the [more] interesting . . . and that’s what I was trying to get at. You can give me any set of numbers and I can frame it in a way that will either damn me to hell or put me up on Mount High, if I really have enough room to wiggle, and that’s the issue, is that, what does it really, really mean?

D: So how do you steer around that, that problem of point of reference? How do you in your leadership, how do you answer these questions of legitimacy and satisfy . . .?

R: I do the calculations right in front of them and I show the minutes, an open book, I seek input from them.

D: So when you say you do the calculations right in front of them, what do you mean?

R: Well count the days off and show all the instructional minutes and do all the math right in front of them so there’s no people saying, “Oh yeah, well you said there’s 487 hours, prove it.” Well no, here it is.

D: Would you ever do the calculations in front of them in terms of student achievement?

R: Never.

D: Why not?

R: Confidentiality . . . the legitimacy of the numbers I present is that I can present this is what we’re going to do about it, and this is why and this again is the outcomes. And it’s my job to answer those questions, I’m the instructional leader of this community cause they come to me and say, “Our kids are scoring 70 percent below average in Reading,” and I say, “That’s right we are, and what should we do about it?” They’re going to say, “What the hell is this guy doing here?” I say first of all, “The time period for when this test, the assessment was done was not the best time, we had two funerals that week. I don’t know if you guys remember this, it was a schedule thing, we didn’t have time to complete it properly” . . . . I have to have the answers and I have to have the answers as to why it is the way it is. And there are times I’ve got to lump it on my shoulders and say you know what, “You’re right, obviously this is something our school is not doing very well so we need to do it better.”

For Rick, in summarizing his values, “it comes down to the balance of accountability versus safety, that you want kids to feel safe so that they can feel that they’re not threatened—you don’t want to embarrass them or feel shame.”
Principals are conscious that they are working in the public realm, where openness is required. Louise believes that test information in the public domain requires principals to be proactive—to be prepared with explanations for results, regardless of whether the averages are positive or negative. “I think we know that with technology and what’s happening regards our public. Check out healthcare these days. No, it’s all transparent. You know, their wait times, their . . . their misses . . . very transparent.” She believes that education should be too, although the results are not going to demonstrate that schools are models of perfection. She acknowledges:

The problem with this is, then when it’s just advertised like if you let the Times Colonist come in and just print off your averages, or if you let the government just print off your averages, or whatever. There’s no explanation. [pause] It can look bad, it can look good, but there’s really no explanation around it. Probably better we print it off and, and, and were transparent on our own.

Principals address this tension between the confidentiality of individual student averages with public classroom and school averages in a variety of ways. Craig has posted several charts and graphs of his school’s achievements in particular areas in the waiting room outside his office—with means and trimmed means illustrated in Cartesian graphs. Pammy has a Data Wall in the school hallway that summarizes the monthly standing of her students. Her school plans and the plans of her elementary teacher groups are public documents, posted electronically by central office.

So I have access to all of my groups [of teachers] and so does the division. So the superintendent of my school looks at every single one of our groups. Like he has eight schools or whatever so he monitors those. So by the next day, we will have, I as an administrator, I would have read all of their documents and their plans. And if I feel there is something there that they [teachers] have not taken into consideration or if I feel they’ve taken on more than they should, or if I don’t think that they have been robust enough or vigorous enough in their attack, I will basically say you need to attend to this, this needs to be looked after because I know at the division level, they’re going to be looking at it as well.

Peter will overrule a teacher’s judgement, if necessary, in the interests of credibility. Whether this data “effect” is a socially constructive or a destructive act in
professionalization will hinge on one’s perspective—a power relation is involved in statistical meaning. Earlier in the year, he “had a very frank and serious discussion with a teacher on staff who did not collect enough data . . . . His class average was way too high for the subject he was teaching and the students he’s teaching, because I teach the same kids in my English class so I know those kids.” Peter felt he had to directly intervene because of the stakes associated with student averages, as well because of his professional sense of being vigilant and as accurate as possible when grading students. “And the second thing is, at parent-teacher interviews, when parents questioning why my marks were so much lower, and why the same kid is doing so much better in the other class [pause] part of it was just for his own credibility [as teacher].” As Peter explains,

So I’m wearing two hats in that conversation with this teacher; I’m wearing the hat of the administrator of the school and I oversee all the kids’ marks, and I’m a colleague of yours teaching the same kids as you are. There’s no way these same kids are getting these kind of marks, so there was the conversation. Well I’ve been at this so long; the good thing is that I’ve been at this for . . . this is my 24th year as principal, so I have credibility I think.

6.8 Leading Questions and Interlocution

When school principals read a display of student averages, they mentally divide students into three groups: student who are at average and those above and below average, and they tend to focus on the below-average students as a potential problem. The below average marks create a doubt or uncertainty that contradicts the principal’s beliefs about appropriate student performance. These doubts generate hypotheses or guesses (usually expressed as questions) about the reasons for the low performance. Questions are an important vehicle for determining the accuracy of the principal’s initial conjecture. Most recognize the fallibility of their initial interpretations, echoing Peirce. As Kirby says, “I could be wrong” and Rick notes that the data display may not yet establish the facts. Principals may ask questions of parents and students, but most usually
start by questioning teachers. No principal has difficulty bringing whole school averages to staff, but they are careful in considering where and how, recognizing that question and data together will influence the self-images of both teachers and students, and shape reality (Swann, Giuliana, & Wegner, 1982) within the school. Most principals begin their questioning with a seemingly casual meeting with a teacher in the doorway of a classroom or a hallway outside the office. Robert asks his questions in a classroom doorway so as “not to step on toes professionally.” He may also ask questions of parents through email, but prefers to phone parents, because “sometimes things get taken out of context.” Because averages serve as “bridges off the island” that is the principalship, it is part of his role to contact parents when needed. In addition, when Robert has questions about below average student marks, he asks a catalyst teacher-librarian to work with middle years teachers, and frequently consults with a literacy consultant about what he calls ED students, those students who are on individualized programs.

D: Robert, would you take that to a meeting of all your staff too, a graph like this [Data Display 3, elementary]?
R: Oh yeah! No doubt.
D: And how would you lead the discussion?
R: I would ask the question, why? Why are levels like this? I mean it wouldn’t be a demeaning question, I would just say, this is the results that we’ve received, what are your perceptions on it, what do you think about? I guess. I would like to, I always want to hear what peoples’ perceptions are because I would have my own, but maybe somebody else’s perception might shake the way I think too.

This amounts to a query about not only the magnitude or scope of the potential problem, but also the problem’s tractability. Moreover, the problem is not only student achievement but, as the comments above reveal, simultaneously one in managing different perspective within the staff. Principals recognize that school size and sample variability have large effects when an average is computed. The range of questions that principals ask is well-illustrated in Rick’s reading of Data Display 3 (elementary):
I’m interested to know how many kids are in the school. What other factors play into their . . . is it a community school? What are their attendance levels like? . . . Do they have behaviour issues? Does their administrator have high expectations? Was there a new administrator coming in? Was there change in teacher and teaching staff? Did four of them go out and four of them come in because that’s reality too? There’s a lot of different factors there. Again those “why” questions, right? Why has it changed? There is improvement in comprehension, which is good, that’s a positive in there. There’s improvement in fluency from one year to the next then it kind of goes down a little bit. You almost wonder if heavy supports are put in place in the district level one year, then they got their scores up, but then the next year they took them away or something because they were showing improvement. Maybe they thought they had the targets in place and then they pulled them away because they wanted to put the reserves somewhere else. They fell off the map a little bit.

We might note that questions are animated by a hunch, by an urge to diagnose, by a search for information that is not evident in the data itself in relation to the average. Principals do not start from a position of complete ignorance, often having teacher anecdotes and other tell-tale signs beyond the statistic at hand. Louise admits that she can be blunt when interviewing some students about their below-average marks: “Are there drugs or alcohol involved here? Is it on the weekends, is it every day? . . . What’s happening here, what’s getting in the way here?” What principals are searching for is explanation, in anticipation of issues that may be raised by parents, central office or community leaders. These questions often are a quest for further information that the principal does not yet have: “What happened here?” “What is going on?” “What is the relationship between attendance and achievement?” “In what skill areas or areas of the curriculum are there a weakness?” “How were marks weighted in the course?” The principal’s aim is to gather more information, while also maintaining or establishing empathy with teacher, student or parent (Landis, 2009). In this sense, the average supports problem identification rather than data-driven decisions, although Louise says that averages “govern student placement” in her school.

As Sr. Margaret reports, how she frames the question is crucial. The questions she most frequently asks of teachers about grade averages are: “Have you spoken to the
student about it? Have you noticed any changes in their performance?” If she has “to get to the point where I say, ‘You know I think you need to . . .’ I try not to use should’s or I’ll say, ‘I’m inviting you to think about this . . . Let me know where you go with it.”

Her sensitivity to the language she uses when posing questions derives in part from experience, in part from awareness that an administrator could abuse his or her power through overly assertive statements, and in part from her religious training in communication and counselling.

When asked about the content of her questions and about the people she is asking questions of, Sr. Margaret reports:

I think it comes down to three; I would be looking at my teacher and saying, “What are you doing...what are you doing to address the issue?” I would be looking at the home life of the child. I’d go back probably and review some documentation if there were any on her, in her file, just to see if there is anything in there. A student that would get a 47 like this here at [our school], the parent would be on the phone, so that would give me an opportunity to ask some questions. I’d look at her attendance; she hasn’t missed a single class, for some of our kids, that’s an issue. And third level would be to speak with the child.

During the interviews, I did not sense that school administrators take notes, or write down their questions, at least during their initial reading of a data display. Interrogation of an average appears initially as a reading tactic, and not as deliberate preparation for eventually engaging others in co-interpretation, nor as a method for surmounting limits in principals’ understanding of statistics. Although these might be motives for some principals, such as when Moira reports that she uses the results “to drive professional discussions,” the questions appear to be an aid in facilitating interpretive understanding. Interrogation of the averages may thus be initially more akin to auto-brainstorming than as the development of questions for teachers to answer.

Most school principals’ commentary reveals an awareness of space when querying the average: whether a student is “up” or “down,” where there is “alignment across,” knowledge of “the right place to pose particular types of questions,” comparing
across disparate classrooms and “geographic locations,” of the demands made by the “hierarchy.” Principals also demonstrate a sense of interpersonal sensitivity that is part of the meaning they attribute to an average: they often know the student or family and where they reside in the community, and understand the personality of the teachers who generate student averages. An underlying structure or cartography for the management of meaning thereby becomes evident, which principals deliberately use to convey to others the amount of authority they should invest in the average.

For example, Louise will call below average students down to the office, because “it ups the ante a bit; they’ve had to sit with the principal.” If the issue is “a social call” which involves some aspect of the students’ personal or social life, she will meet in an informal lounge down the hall from her office. But if “it’s more of an academic call, we’re in my office” where she has her computer and Power School software. There, she can “click on a mark, you can go inside, see all the assignments they did do, didn’t do, what they got in a mark, you know.” Inquiries may be about attendance, and other issues requiring the services of a chaplain or counsellor, whose offices are next door. Similarly, Sr. Margaret carefully chooses the setting in which she poses questions to teachers about averages: she never questions a teacher’s marks unless she and the teacher are alone. In her “continuum of intervention,” the lowest level is an informal chat at the main office counter, where she asks probing or leading questions. More formal meetings around an average, called a Conference, occur in a meeting room adjacent to the office. There is also what might be called a Conference Plus that would take place in her office on the rare occasions when this third level of intervention is necessary. A fourth level is what she calls a Super Conference that she convenes with her at the head of the table in the school’s board room. The Super Conference would involve herself, the teacher, student and parent, and Sr. Margaret recognizes this as a high pressure meeting for the student.
In the privacy of her office, which she deems a sanctuary, Sr. Margaret sometimes hears students’ concerns about the marks they have received, and she usually asks students to address the problem directly with their teachers. However, she will help students prepare for discussions about marks with teachers. “I have a policy, not a policy; I have a practice or a philosophy. I will ‘prepare you for battle but I will not do the battle’ in these matters.” Her practice of using her office as a preparation room applies equally as well to teachers if they have difficulty with parents over marks.

Intervention, interrogation, and investigation may thus be terms that are too connotatively forceful when describing school principals’ actions with the average: inquiry and interlocution better describe their styles when ascertaining and shaping the practical meaning of an average. The statistical representation will frequently prompt the principal to delve more deeply into how the student’s mark was calculated by the teacher: the types of evidence which were assembled; the weighting attached to each piece of evidence; or the range of assignments which compose the average. It is rare for a principal to directly override the mark assigned by a teacher. Kirby has overridden teachers’ marks twice in his career, and Peter has directly changed teachers’ marks only a half dozen times during a long career in administration. Phillip is explicit that he sees questioning as the first step on the ladder by which he expresses his authority: his next step is “making a statement.” In seventeen years of dealing with marks, Sr. Margaret estimates that she has only intervened to directly overturn a teacher’s marks a handful of times. Principals are aware that their questions will accomplish the professional moderation or substantiation that is sometimes required. This would indicate that principals’ questions are a rhetorical device—an indirect strategy for asking teachers to reconsider marks that may be too high, too low or unfair to particular students.

Yet underlining many principal’s questions is the urge to justify, the need to generate answers to questions that may be asked by others down the road. In that sense,
the principal’s questions are those they anticipate will be asked by others who read the same data display. However, when asked, principals indicate their questions are not the imagined voices of another, virtual reader who might be looking over their shoulder, to whom they feel accountable\textsuperscript{12}. These are their own questions. Notwithstanding this perception, the questions do not seem to derive from a lack of operational knowledge, either in mathematics or in pedagogical processes, but rather from an inability to explain student performance or the statistical author’s computation to an anticipated audience. The students’ overall pattern of averages must make sense to the principal. Nevertheless, the average spurs a hunch or initial diagnosis without excessively limiting the range of possibilities. Therefore, the average as notation does not inevitably predetermine a conclusion, but rather both opens and delimits a zone for further inquiry.

Several features are evident in the questions principals ask about averages. First, the question is often of a first-order, journalistic kind, implying or pointing toward a hunch or preliminary diagnosis. Second, many are closed questions, indicating that an issue has arisen which calls for a position to be taken and an explanation or justification to be provided. Third, many verbs are in the conditional or subjunctive present tense, even though the data may derive from assessments a month or more before. Even though the statistic may reflect a past circumstance, answers are required in the present. And fourth, principals are uniformly precise about to whom they would look for answers, within what Pammy has called the educators’ “circle of influence.”

\textsuperscript{12} Peirce rejected the homunculus argument, the philosophic fallacy which arises when we explain some phenomenon in terms of the phenomenon itself that it is supposed to explain (CP2. 435-436)—in this case, explaining principals’ questioning by attributing it to others’ questioning. One of Peirce’s papers which has baffled some scholars of pragmatism, \textit{A Neglected Argument for the Reality of God} (EP2. 432-435), may be read as a refutation of the homunculus argument.
According to Pammy, any test result is a reflection of the school, and not just of an individual teacher’s work, so when results are received, the staff will celebrate or commiserate as a group. When she gets questions about the accuracy of results from teachers, who may argue for more resources, who may attribute student averages to socioeconomic conditions, or who may complain about parents not helping with homework, she takes the group back to teachers’ “circle of influence,” which is within classrooms and schools. “Whether parents read with their kids at home at night is outside our circle of influence, whether kids get to bed on time, or they’re adequately nourished, or if they have homework time, that’s all outside.” She implies that her teachers should stop blaming the families, the community, or other teachers for students’ low marks, focusing instead on those matters they can control. The student’s experiences in the classroom and school are within a teacher’s circle of influence, and should be the absolute best they can be.

Implicit in Pammy’s commentary is the idea that the meaning of a numeric figure is determined within a school, and that people outside the school are prone to misinterpret the significance of a percentage average. Pammy believes it is the criterion scales that should be transported into households, not the student mark as an average. She is interested in outcome-based education, where teachers use formative assessment to determine whether children are meeting, progressing toward, or excelling at a particular outcome, or whether there is a lack of evidence to demonstrate the child’s learning:

And so rather than using numbers to describe the learning we now use excellent, meeting, limited, or no evidence. And so up to Grade 6 in our schools, there are no numbers given; Grade 7 and 8—we still give numbers only because and that’s what parents are expecting. It’s certainly not what we feel necessarily that is the best, that gives the best indicator of whether children are learning the outcomes. But it has to be made into a number, still at Grade 7 and 8.

For Rick, anomalies in grades, as deviations from an overall pattern, cue “why” questions which will most often send him initially to the student. He typically seeks
further information from students about their home situation, about their attitudes and feelings, and about their circumstances in general, before approaching the teacher.

According to Rick, if he makes assertions about the cause of a low average, he may create “paranoia” among teachers whose contracts are renewed every two or three years in an on-reserve school. Invariably, he also asks questions of parents too, rather than making pronouncements or offering up direct conjectures because strong assertions cause parents to become defensive. Statements such as “I noticed Joan missed . . . .” or “I noted that Ralph did not do well on . . . .” as an observation with evidence to open the conversation, with subsequent use of the first person plural “we” is used as a way of “hooking parents into the conversation.” Similarly, Sr. Margaret says that asking questions about a student’s average “opens lots of doors for you.” She sometimes has a student who is performing “at max” and suddenly “plummets;” so she begins to ask questions that lead her to difficulties at home. “As I tell my teachers, ‘We never know what goes on in the home the night before.’”

According to Phillip, for many elementary teachers, having to assign a number grade is an impediment to a developmental approach in pedagogy. Communicating to parents with averages does not convey whether the child is performing at grade level, and does not indicate what the child’s particular strengths and weaknesses are. In some ways, the number becomes a label and distracts teachers from communicating more fully with both parents and students about students’ learning. “I think that it can be if we get so focused on that number, then we’re in trouble,” Phillip believes.

It remains unclear whether school administrators’ queries express the hunch or conjecture itself, or whether the principal’s question is separate from but complementary to their conjecture. Close scrutiny of school principals’ commentary suggests the two may be fused in administrators’ practice. If so, then the informal inquiry of school principals may differ from formal methods of scientific inquiry. In formal inquiry, the
interrogative suggests a general orientation to a problem, and a distinct hypothesis formally suggests answers one would probabilistically find through induction. On the other hand, within the informal inquiry of a school principal, the hunch or conjecture is often implicit within the question, thereby sustaining the subjectivity and hence open-mindedness of the inquirer when seeking an explanation (Lipton, 2004). As one scholar (Ilie, 1994) concluded—after analyzing courtroom, media talk show, and parliamentary questions—rhetorical questions have two pragmatic functions: those making the query may not be asking but instead conveying an underlying statement more subtly; and the query may elicit diverse responses from an interlocutor which may range from information to action to counter-argument to reflection. The questioner’s intentions will shape the response.

However, much remains obscure in principals’ commentary about this non-linear route from (a) interpretation to (b) inquisition to (c) interlocution. Whereas interpretation and inquisition are not serially distinct but intertwined in school principals’ readings of statistics, principals’ initial queries of the data display are pre-cursive to interlocution with others, but not necessarily recursive. They do not necessarily return and re-read the statistic in light of what teachers or others say. Thus, much remains to be clarified. How do principals select or refine their initial conjectures as questions for others? What are their specific intentions when posing a question of others? What rhetorical style do they adopt when posing the question? What specific expectations might accompany the question? How do principals actually manage the ensuing interactions with various constituencies? Pursuing these lines of inquiry is beyond my purposes here. But it is abundantly clear in principals’ testimony that the average as Sign has been transformed during statistical reading from inferential proposition to interrogative proposal. What was a conclusive result for the statistical author’s inquiry has become a question at the beginning of the statistical reader’s own inquiry. The journalist Christopher Hitchens has
offered, as a maxim, that what can be asserted without evidence can also be dismissed without evidence (Hitchens, 2007). School principals do not dismiss the evidence and do not generally assert with statistics, but rather interrogate an average’s meaning in consort with others.

6.9 Consequents and Antecedents

The answers to the questions that principals ask after reading an average are often sought retroductively, moving diagnostically from the average as consequent to antecedent. Although customarily seen as the logical act of drawing conclusions from known premises, what is called an inference may also embrace an equally rational form of thinking which opens new avenues of inquiry. In other words, the average is provisionally held as a consequent, and the author infers backward to unknown premises. In school principals’ abductive line of inferencing, the statistic precedes and guides the generation of a hypothesis or conjecture, as compared to formal quantitative inquiry, where the hypothesis precedes and guides the generation of a statistic. These reasoning processes do not so much supply the rationale for action, as Dewey (1941) claimed for warranted assertibility, as reveal how such a rationale is created. Therefore, we can hardly say that thinking with an average is simply an act of hyper-rationalization, invoking a rigorous search for rules through inductive reasoning or an ironclad deductive inference from which conclusions are drawn. Rather, the average as representation serves as a point of departure in thinking, as an initiating prompt for a line of inquiry where the conclusion is provisionally held to reconsider anterior conditions or premises. For as Peirce argued, the logic of practice is synonymous with abductive reasoning (EP2. 226-241). The features of this mode of reasoning predominate in school principals’ readings of data.
For Robert and Sr. Margaret the answers to the questions that arise after reading an average could be sought in the student files, as a kind of bio-historical inquiry, and they might ask a teacher or learning consultant to look at the files.

D: So how would you go about finding out the answer to that question? Why?

R: Look at his “cum” file [cumulative]. Look back into his “cum” file, is there a referral to student services? Is there an Ed Psych report? Is there a Woodcock Johnson done on the kid? What other math assessments have been done on that student in the past? Has there been a standardized test done every single year? [pause] What comments have been written on report cards by teachers? And there’s a ton of investigating that you can do with a student like that. You could have a conversation with parents about past history. Maybe this kid moved around a bunch when he was in grades say 1 to 5. It doesn’t say anything about Mum and Dad being together. Maybe he moved around a bunch and that was just something that didn’t resonate with him.

D: Right.

R: So many questions that you could ask. Maybe he just does have a learning difficulty?

Rick is more methodical in pursuing answers to his questions. Rick likens the search to a parlour game where he must use a process of elimination to find the solution:

D: So we come back to this word “factual,” are there any facts in this graphic?

R: Nope.

D: Why not?

R: Well there are facts, sure there are, but are they useful facts? You ever play CLUE?

D: Mm Hm.

R: . . . a game where you make suggestions, there’s a lot of stuff on that, a lot of different rooms and things, there’s a system to it, though at least in our family there is. I devised it, my wife accused me of cheating, my daughter liked it so much she uses it every time now.

D: What’s that system?

R: Process of elimination. All those things don’t mean anything, until you can have a reference point. So you’ve got three things in a clue: you’ve got weapon, person, place, right? Until you know one is right, you can’t effectively try and find out all three without starting with one . . . whereas if I
can narrow my focus, and say this is what I’m looking for, it may take me a couple of turns but eventually I’ll know what’s in the parlour . . . Then people will look at me like you’re crazy because you got hit. Of course by then they’ll figure out but by then I have this gleam in my eye, I can work my way through it.

Both Rick and Kirby, who are in very small schools, would search for answers in the original test questions. Rick would review each test, student by student, looking for shortcomings and especially for deficiencies in particular curriculum areas:

Then you can come up with a common theme, you know like maybe 10 kids did poorly because of their reading, they didn’t understand the question. OK, then we know where we need to go. Maybe the others did wrong because they got lazy and skipped steps you know; we know what we need to do, you can group them and start to build up their strengths.

His conjectures are the starting point, which may or may not be eventually confirmed inductively. Rick’s detailed reading of Data Display 3 illustrates the entire trajectory of interpretation, taking him from the result as consequent to hunches to interlocution:

R: OK, first of all we’re testing at reading, it’s showing me comprehension, fluency and a total reading score which I’m assuming is going to be some sort of average or a combination of the two scores together to give me that. So I like that right away because it’s giving me something specific. Now there’s one thing missing from all these things that I would really consider to be important, and that would be the number of students tested.

D: Why is that important?

R: We test three kids, and Bobby decides not to try, so I have only two, my mean score is going to drop by a third before we’ve even started. The greater the number of students, the least, the little more effect each student does end up having on the score, you know what I mean.

D: Right.

R: I don’t really know what the 360s, and all that sort of, oh I guess maybe Grade 3 and that would be level three. Red is the district, blue is the …these are relatively consistent scores throughout the three years. I mean there’s growth, there’s a little regression in the third, it shows that according to this that the students aren’t quite meeting the average in the whole district but then again number of kids tested could be a factor and this is [the] number of schools brought into it. I’m interpreting this at the grade level 3; that this is like a 3.1, 3.2, 3.3 so we’re going by one tenth, maybe one eight break up. And there’s about pretty consistent spread of an eighth for the
comprehension. For fluency there’s a slight deviation about the same, they’re relatively within a tenth consistent across the board. I guess what this tells me there’s been little change in three years. There’s been a little change but not a great deal.

D: OK, and if you had a graphic, a data display like this that came from [the local First Nations education authority]?

R: Yeah, we have it.

D: How would you proceed?

R: For me, why don’t you sit down and talk with everybody as to what’s going on? Well we’ve got a teacher who for three straight years has been doing a little bit below what the rest of the school is doing, is it her or the students? If it’s one year, it might be just the students, if it’s three straight years, it’s not just the students, it’s a bigger systemic issue which could be flowing from Grade 1 and 2 so I would probably want to do. I would want to look at the Grade 1, 2, 3 teachers and have them have a conversation with somebody whose kids are performing at a higher level and say, “What are you doing that we’re not?” “What’s different about our kids than not?” It has to be a conversation that’s professional.

For Phillip, averages are used as a form of risk appraisal and management for below average students: however, the risks are not necessarily computed in terms of the proportion of the student population who are scoring below average or hand-calculated from a student’s report card. Instead, Phillip, like Robert sees the teacher-assigned mark as a barometer of whether a particular student is “at risk” of not succeeding, or whether the school overall is at risk of being categorized as below-average in comparison with other schools. Such a view impels him to mitigate the risk as best he can, whether in pre-existing teacher-student relationships or parent-child relationships.

When reading Data Display 3, Pammy too draws retroductive inferences from the average, drawing the provisional conclusion that a problem exists, but seeking an explanation at earlier grade levels:

Well there’s not enough growth, we would need to put . . . obviously in my opinion they’ve either had a plan in place for reading that they’ve worked at some, but not to the extent that they’ve should. So we need to dig deeper, why our students are not as fluent as they should be. The total reading scores all the way around are lower than the district. But I mean again these are three different groups of kids, we need to go back and see what they’re doing in Grade 1 and Grade 2 and seeing where their skills are. It’s not just a Grade 3 problem; it’s a Kindergarten, Grade 1, 2, 3 problem. So in my school we celebrate together and we cry together,
is what we say, “So what are you going to do? What are you doing in kindergarten? Where are your kids?” We need to assess, assess, assess—intervention, intervention, intervention. And the same in Grade 1 and the same in Grade 2.

For Moira as well, below average student performance would cue a variety of actions to remediate anterior conditions, including calling in additional resources for particular students such as instructional therapists, and speech and language pathologists. She, unlike other administrators, would phone and ask for more resources from central office. Louise does indicate that demonstrations of poor performance, or the evidence of results as represented in high school credit accumulation, are sometimes used obliquely in meetings with central office staff to request additional resources. In that sense, below average student achievement does come into play, Louise admits sotto voce in the interview, but not in a formal way. Saskatchewan school administrators do not consider average output as coinage to formally barter or negotiate inputs with central office.

Reasoning about the average as consequent toward antecedent conditions is different from reasoning from the average as cause toward effecting particular types of changes in student learning, in instructional arrangements, or in teachers’ practice. Whereas a consequent is a logical term denoting a conclusion drawn from premises, the latter is prospective in imputing causal relations between categories of events (the relata) for making predictions. The principals’ agency is manifest in the former, but not especially discernible in the latter. What is now recognized as an axiom in formal propositional logic as Peirce’s Law (the non-intuititionally valid formula: the antecedent A. must be true if you can show that the Consequent necessitates A. to be true) may be regarded by some researchers as an especially esoteric proof in mathematics or syntax application in computer software design and cryptography (Nievergelt, 2002). Peirce’s Law has enabled computer programmers to mathematically work around the law of the excluded middle (the binary notion that a proposition is true/not true), linking game
theory with ordinary propositional calculus. Yet Peirce’s Law may have practical importance because it enables leaders to generate solutions to student achievement problems. Abductive reasoning is a form of logic that encompasses heuristics and is integral to school principals’ interpretations of average student achievement.

In brief, the conjecture (1c.) about school administrators’ defining consequents in relation to classroom assessment results rather than large-scale assessment is not entirely corroborated. School principals are ready to interrogate any set of averages, regardless of their source. Moreover, principals do not frame consequents as synonymous with prospective effects, in the sense of what the average causes in a temporal sequence. Rather, for school principals, the average in multiple forms is held as a consequent, which sparks a retroductive search for information about conditions at earlier grade levels, within the school, or within the home which would explain the result (Hintikka, 1998).

6.10 Causes and Effects

When asked about effects of the average, Saskatchewan school administrators often frame causal sequences within school cycles, thus rendering it nearly impossible to extricate the principal’s sense of agency from broader organizational rhythms. For most principals, the direct contemplation of averages is associated with the authorization of report card marks. Principals’ habits in interpreting averages thus are often indistinguishable from the administrative routines they have established to order school activity, or from those larger administrative cycles that govern public education. The principals in this study often position themselves in relation to the average not as causal agents, but rather in relation to an individual student and that student’s trajectory through schools, hoping to elevate student learning within a larger educational cycle. Yet for most of these principals, the statistical average as construction means much less than the sum of its parts. That is, student counts, per-student allocation of dollars, and class size
are considered to be more “real” and have greater significance than average student marks. Often, principals’ administrative practice is governed more by financial inputs than by student outputs.

Thus principals periodically consider outcomes data, varying by the different reporting intervals in elementary and secondary schools. For Pammy and Moira in elementary schools, reviews are monthly, whereas for secondary teachers, reviews are usually conducted quarterly. Some principals, such as Peter, sit as final causal agent “at the end of the rainbow,” receiving teacher-assigned grades as idiosyncratic averages and taking seriously the responsibility for reading and authorizing report cards with a signature. As Peter states, the averages can represent “a pot of gold” in the form of scholarships over which he presides. Other principals, such as Pammy, adopt a monitoring role and delegate responsibility for the process of interpreting student marks to teacher teams and the vice-principal at monthly staff meetings.

As evident in Table 6.2, for most principals, average student achievement becomes germane within the ongoing redistribution of resources in schools; effects are described in terms of that cycle, which thereby position the principal in different ways in relation to average student achievement. Public school principals described a cascade of administrative decisions each year: each decision involves consideration of averages in ratio form. In the principal’s view, the central animating decision is often the annual provincial budget which sets grants that are distributed to school divisions. Once these grants are received, central office officials and the school board set pupil-teacher ratios (PTR) that are of prime importance in determining the staffing allocation for each school. Once set, the PTR determines how many teachers will be available in the school division, and thereby predetermines average class size within the division.
A four-stage process is thereby triggered at the school level. First, the PTR is set at
the board office for the succeeding school year, a decision in which principals have
varying degrees of influence. Second, principals then calculate their staff complement,
determine course and program offerings, and assign students to particular classrooms.
With knowledge of average student achievements (in terms of report card grades and
promotion decisions) from the previous year, the school principal then matches students
and teachers and assigns both to various classrooms, timetabling varying class sizes and
mixtures of student need and available teacher expertise to maximize student achievement.
Third, “and in theory” as Phillip explains, this leads to varying levels of opportunity for
individualized or differentiated instruction by the teacher. Fourth, principals “hope” this
individualized instruction will in turn lead to better student achievement in the ensuing
school year. If the first two steps are the school principals’ prerogative, the latter two steps
are conceived as the teacher’s and student’s responsibilities.

The entire process is repeated each school year and the output from the previous
year feeds forward into decision making for the succeeding year in a causal sequence.
In that sense, the students’ marks as signs serve as a sorting mechanism, not for students
per se, but for the dedication of professional resources to particular groups of students.
Indeed, for Craig, the role of the principal itself can be conceived as a “numbers game,” by
which he means an ongoing set of administrative accommodations and adjustments when
responding to the many pressures created by various averages of input and output, and
their computation. He prefers averages to remain “grey” and “vague” in their calculation
and meaning, because that obscurity is where he finds latitude for maneuvering. The
management of meaning thus involves balancing the significance assigned to several
averages, not only by teachers, but also by central office personnel in their decision-
making.
Table 6.2

*Causal Chains, Perceived Effects and Agentic Position, Saskatchewan School Principals*

<table>
<thead>
<tr>
<th>Respondent</th>
<th>Causal Sequence</th>
<th>Perceived Effects</th>
<th>Principals’ Agentic Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Craig Tremaine</td>
<td>PTR &gt; allocation of resources &gt; differentiation of instruction &gt; average student performance</td>
<td>Average is flexible point for ordering students for comparative purposes</td>
<td>Principal uses average as conversation starter to moderate expectations</td>
</tr>
<tr>
<td>Kirby Klein</td>
<td>Average achievement &gt; inappropriate board and ministry conclusions &gt; community withering</td>
<td>Averages have misleading and attritional effect on school viability</td>
<td>Principal is victim of inappropriate central office or ministry decision making with averages</td>
</tr>
<tr>
<td>Louise Simpson</td>
<td>Average &gt; sorting above/below average in watch list &gt; call in additional resources and student supports</td>
<td>Average enables individualized attention for specific students</td>
<td>Principal triages students and calls in additional resources for below average students</td>
</tr>
<tr>
<td>Moira Gullickson</td>
<td>Average &gt; reorganization of student groupings &gt; request for additional resources &gt; greater staff focus &gt; above average achievement</td>
<td>Averages are relative standard for determining success or failure in instruction</td>
<td>Principal drives professional discussion with averages to focus on and improve achievement</td>
</tr>
<tr>
<td>Pammy Reid</td>
<td>Average &gt; parental misunderstandings &gt; simplified and non-evidence based instructional approaches &gt; lack of criterion improvement in outcomes</td>
<td>Averages perpetuate traditional and outmoded lines of evaluation which impede clear communication of curriculum criteria to students and parents</td>
<td>Principal resists reporting and discussion of average student achievement as out of date and inappropriate at elementary and high school levels</td>
</tr>
<tr>
<td>Peter Getz</td>
<td>PTR &gt; redistribution of resources &gt; credit acquisition &gt; above/below average sorting &gt; post-secondary education or direct entry to workforce</td>
<td>Averages determine a student’s trajectory during and after high school</td>
<td>Principal is balance point in scale of educational justice who reallocates available resources to enable students to fulfill their academic potential</td>
</tr>
<tr>
<td>Phillip Brunswick</td>
<td>PTR &gt; redistribution of resources &gt; smaller PTR for below average &gt; greater differentiation in instruction &gt; improved student marks</td>
<td>Average supplies useful information for professional and administrative reflection, but increasing concerns about public misuse for accountability</td>
<td>Principal reflects school division goals and balances resources in a moderating way as part of systemic cycles</td>
</tr>
<tr>
<td>Rick Gabriel</td>
<td>Societal obligation to grade &gt; assignment of average/above average &gt; differential flow through curriculum</td>
<td>Averages are part of educators’ stagecraft for legitimizing their professional or administrative position of authority</td>
<td>Principal uses average as clue to seek explanations for anomalies in student marks to regulate flow</td>
</tr>
<tr>
<td>Robert Hanson</td>
<td>Below average &gt; discussion with teachers &gt; greater staff focus &gt; improved student performance</td>
<td>Averages focus attention on those students who do not match the norm</td>
<td>Principal uses averages as cue or red flag for identifying individual student problems</td>
</tr>
<tr>
<td>Sr. Margaret Westlake</td>
<td>Average achievement &gt; labeling of student &gt; educational stigma &gt; disempowered students who do not fulfill potential, or below average &gt; direct intervention of four kinds depending on severity &gt; refocused staff and parental or student attention on academics</td>
<td>Averages both signal need for and impede direct exercise of educator will</td>
<td>Principal along with staff are direct interveners in students’ academic trajectory</td>
</tr>
</tbody>
</table>

*Note:* ➞ leads to

PTR = Pupil-teacher ratio

---

235
For all these principals, moral propriety is integral to the “proper significate outcome,” as one of Peirce’s definitions for significance (CP5. 473; EP2.316-320; EP2. 496-499) in the final interpretant, or the practical meaning of average student achievement. This sense of moral purpose, particularly that between the teacher and student, informs school principals’ readings of the average, and their notions of final or ultimate causation—where the whole brings out the parts. Final causation, as an overarching sense of purpose governing the day-to-day or efficient causal decisions, is thus integral to school principals’ sense of propriety. While Louise does not know whether “average” represents “good enough” or “unacceptable” performance, which would enable her to use it unequivocally to accomplish an initial triage, she, like all the other principals, is clear that “below average” is an element of sorting. Moira too believes, as a principal, that “average” may signal acceptable achievement to some parents and students, but as a parent of three, she believes that “average” performance is “unacceptable” and is thus actionable. Peter is clear that he will not alter averages for a report card grade after they are entered and computed within a district software package, considering this to be a “manipulation.” Rick’s ethical frame of reference extends to his approach to reading a data display itself, and not just his conception of what the statistic means: he deliberately does not look at individual student names when initially scanning data. This, he believes, permits him to be objective when identifying those cases that he feels require his attention, forestalling claims of bias. Indeed, many principals criticized the provision of background information for individual students in Data Display 1, as inappropriate when interpreting student marks, seeing it as detracting from an unbiased appraisal. Kirby’s skepticism about averages derives not only from his doubts about the
algorithmic applicability of a single formula to the diversity and variation he sees everywhere in social phenomena, but also from the professionally unconscionable act of not covering a curriculum: teacher marks are not comparable between rural and urban Saskatchewan in his view because accredited teachers ignore substantial segments of the curriculum when teaching, so that both the denominator and the numerator are non-equivalent for comparative purposes.

Sr. Margaret is guided by a strong moral compass that overrides her concern with the symbolic label of “average” that might be attached to a student. “Below average” student achievement is not only a slight to her and her school’s reputation, which are synonymous in her mind, but also an affront to her moral stance that it is unacceptable for any person in her school to not seek and reach some form of excellence. Craig’s “grey zone” in large-part derives from the subjective and moral ambiguities in grading as an act of student evaluation: he, like Phillip, is convinced that the average as centre must encompass a sufficiently broad zone so as to give every student the benefit of the doubt, especially at the threshold of failure, but also at other thresholds including identification of exemplary students who will be recognized for awards. Assigning student grades is especially subjective in the humanities, school principals believe; the many difficulties in defining a student’s student achievement stem from the process of evaluation, which is not only the assignment of worth, but also a human and not God-like choice which will shape a student’s future within school and beyond. In Peter’s demonstration of the idea of the mean in student achievement, as photographed in Figure 6.5, he indicates that he must annually balance a host of administrative choices including course offerings with post-secondary aspirations, parental expectations, and yearly budget allocations, not just student marks.
Such choices, school principals insist, are more than problem-identification. Most principals underline that averages are used in making school-based choices: “They inform decisions, your numbers—I mean data informs decisions,” Robert argues. “It tells you if you’re moving forward, what you’re going to do with a student. If they’re achieving really, really high, then you need to move them above the high, or [they fall] on the other side.” Similarly, Louise is adamant that averages govern program placement, pointing to spring meetings of staff from several middle years schools who review student performance and make recommendations for program placement before students make a transition into high school. Both Craig and Peter are explicit that promotion and graduation decisions hinge on student marks. However, these are not the clear-cut, dichotomous decisions as formalized in decision theory, but rather the
piecemeal decisions that accumulate over time in individual classrooms, in teacher committees, in various teachers’ grading decisions, or in teaching assignment and timetable choices made by principals. Some principals will point to the student’s cumulative folder as the extended record of such accretive decision-making. Accordingly, Saskatchewan school principals incarnate Peirce’s provocative idea that “interpretation involves a power of the symbol to cause a real fact” (EP2. 322-324).¹³

One difficulty here is that moving relatively small numbers of students from the “below average” category to the “average” category alone will not arithmetically shift a classroom’s or a school’s overall average substantially in relation to others. To improve a classroom or school average requires elevating the performance of large numbers of students who achieve at average. Pammy recognizes this problem, and it underlies her critique of the average as meaningless for communicating pedagogically-useful information to parents. Far better to use criterion referencing, she believes, with evidence of student attainments for particular curriculum outcomes rather than mislead parents with ever-shifting averages. The second difficulty is that if averages generated exogenously through a large-scale assessment arrive on their desk, principals do not have singular authority to manage the meaning of these averages, except in comparison to classroom averages. Averages derived from large-scale assessments may be readily dismissed by the principal and teachers as a snapshot, as a single instance of student performance, but they are not seen in this way by the employer or the wider community, who attach considerable importance to these marks.

¹³Peirce also noted that because a symbol must be interpreted, it necessarily reproduces itself, but that “we necessarily defeat ourselves if we insist upon reducing everything to a norm which renders everything that happens, essentially and ipso facto, unintelligible” (EP2. 322). Peirce was explicit that when communicating an interpretation, the interpreter becomes a “man-symbol” who is “essentially a purpose, that is to say a representation that seeks to make itself definite,” thereby determining judgement and assertion (EP2. 323-324).
In short, the conjecture (1d.) that school principals will discount the significance of an average as involving the comparison of two means is not supported by principals’ commentary. Principals will actively compare students’ report card marks with their other marks, with the overall mean generated within a classroom, and with marks generated by a variety of measures exogenous to the school. However, average budget allocations, average pupil-teacher ratios, and average class size are considered to be the most important averages which ultimately have the greatest causal influence on student achievement.

School principals’ reasoning patterns with student achievement statistics recall William James’ (1890) apt metaphor when describing the stream of consciousness intrinsic to interpretation:

> Let anyone try to cut a thought across the middle and get a look at its section, and he will see . . . the rush of the thought is always so headlong that it almost always brings us up at the conclusion before we can arrest it. [Introspective analysis] is in fact like seizing a spinning top to catch its motion . . . (p.244).¹⁴

Principals are in many ways that spinning top, and in their contemplated actions with the statistical average, they personify an ever-oscillating axis around which student evaluation practice occurs. The sign is amenable to principals’ influence, but they generally prefer to ask questions and express their authority indirectly through conversation or through discussion within a “circle of influence.” In short, the average as statistical phenomenon is a point of balance first for students, second for teachers in their classroom assessment practices, and third for the community of parents. Principals are reluctant to intervene with teachers to moderate the effects of the second, but have no

---

¹⁴ Peirce disputed this claim by James (CP8. 90), insisting that thought can be mathematically depicted in quantities, dissected and inspected without introspection. Many of Peirce’s criticisms of overly psychological views of interpretation and perception can be found in his review (CP8. 55-90) of James’ (1890) foundational text, *Principles of Psychology*, and in an unpublished article (EP2. 371-397). Although Peirce acknowledged empirical psychology’s accomplishments as an observational science, he feared that psychology’s failure to consider specific forms of reasoning portended “an era of intellectual epilepsy” (EP2. 471).
difficulty mediating the first and third. Averages depicting student achievement enter the school principal’s office as one of many statistical statements, but often leave the same office in the form of questions about students who apparently do not reach the average’s fluctuating threshold. What may be a product and proof terminating the statistical author’s inquiry, becomes a sign which sparks the statistical reader-as-leader’s own forms of inquiry.

Consideration of student outputs is only one facet of the principal’s role; the interpretation of statistics is a recurrent and substantive aspect of that function. Essentially, school principals conceive of average student achievement as a dynamic point of balance. They consider the sign periodically for the individual student’s report card averages, episodically with teachers in their appraisal of student performance across a classroom or grade level, and more frequently for the entire school in considering its various outcomes as evaluated through large-scale assessments which originate from outside the school. Animated by a sense of moral purpose, Saskatchewan school principals would habitually manage the diverse meanings of these three interacting types of averages by asking leading questions (Marquardt, 2005) when the statistic signals disequilibrium.

### 6.11 Recapitulation

Based on interviews with 10 school principals in Saskatchewan, I adduce that:

1. School principals’ conception of average student achievement is synonymous with a centre of balance or point of equilibrium in student, classroom and school affairs.

2. School principals hold multiple and nuanced conceptions for average student achievement, but would not express their conceptions in terms of various statistical forms or mathematical formulae.

3. Viewed pragmatically and semiotically, the average is both an index and a symbol which is subject to many and often, nearly concurrently-held, meanings for school principals. Because average student achievement as a statistical icon does not resemble the student, principals would render it more
meaningful through metaphor, colour, and gesture so that it has indexical qualities.

4. In its received form, the statistical average would be perceived by principals as a cue or clue which prompts further inquiry into antecedent conditions for those students who have below-average performance, or whose performance is declining.

5. Different types of measures of central location as a statistical concept would be considered when interpreting data, depending on the principal’s purposes. The average would be considered as a weighted arithmetic mean for individual student marks, as a mid-range or modal distribution when considering classroom averages, and as the arithmetic mean when considering whole school averages.

6. When principals read a display of student averages, they would triage students into three groups: those who are at-average and those who are performing above- and below-average. This allows them to focus their attention on students who are not achieving well. This triage is not exogenously imposed on principals when they read statistics, but rather is endogenous to the interpretation of data itself as a reading tactic.

7. After principals read a statistical average, they would often transform it from a proposition to a rhetorical question they would subsequently take to various audiences for answers. Principals’ conjectures or hunches are implicit in the questions they ask.

8. Abductive reasoning with hunches and provisional diagnoses toward the best explanation predominates over drawing deductive conclusions or an inductive search for rules when principals are initially interpreting an average. Principals provisionally hold the average as consequent, as the starting point for a search into antecedent conditions.

9. Principals would use the average pragmatically as a fulcrum to lever a variety of instructional or home-based changes by asking questions about those students whose achievement is deemed below-average.

10. The agentic or efficient causal thinking of principals in relation to average student achievement, and their habitual readings of data, are inextricably bound up with planning and reporting cycles, cycles which principals lead and in many cases have structured and institutionalized themselves within schools.

11. The final causal or effective thinking of principals is expressed in their sense of moral propriety about averages, particularly for those students perceived as performing below-average.
We generally know when we wish to ask a question and when we wish to pronounce a judgment, for there is a dissimilarity between the sensation of doubting and that of believing.

Charles Sanders Peirce

Chapter 7: Effects of Framing Average Student Achievement on School Administrators’ Problem-Solving Styles

In this chapter, the effects of perceptual judgment on Saskatchewan school administrators’ conceptions for the average are gauged, depending on whether an average is framed positively or negatively, and whether the average is made public or held confidential to the school.

“The elements of every concept enter into logical thought at the gate of perception and make their exit at the gate of purposive action,” Peirce argued, “and whatever cannot show its passports at both those two gates is to be arrested as unauthorized by reason” (CP5. 212). He was adamant that (1) the meaning of any representation in any kind of cognition—virtual, symbolic or otherwise—first arises in the senses; (2) perceptual judgements contain general elements; 15 and (3) abductive inference shades into perceptual judgement without any sharp distinction between the two (CP5.180-212). All conceptions have as starting point or first premise, a perceptual judgement (EP2. 223-233). Peirce was among the first to conduct randomized experiments about such judgments, using Fechner’s theory of psychophysics (Peirce & Jastrow, 1885). Therefore, we might look to perception as shaping school administrators’ conceptions of the average. If most administrators conceive of the statistical average as a centre-of-

15 Peirce did not specify in his use of the term “general”, whether he means “generalizable”, “universal”, “abstractive”, or “habit-forming”.
be thrown out of balance? That is, what factors will significantly disrupt their beliefs about average school achievement?

More precisely, my research question is:

**How does framing “average” student achievement, in the positive and as a statistic made public, influence school administrators’ conceptions of the average and their problem-solving approaches?**

The associated hypotheses are:

a. School administrators will adopt more rational and analytical problem-solving dispositions when school averages are framed negatively than when they are framed positively.

b. School administrators will adopt more experiential and intuitive problem-solving approaches when averages are framed as confidential to the school, than when they are framed as in the public domain.

c. Vice-principals will have more rational and analytical pre-dispositions to the interpretation of average school performance than will principals.

d. Experience as an educator and as an administrator, position, gender, education level and statistical self-efficacy will not substantially predict school administrators’ problem-solving pre-dispositions when reading statistical averages.

Many recent studies demonstrate that negatively-framed situations and information receive more emphasis in human cognition, emotion, and attention than positively-framed ones of comparable magnitude (Kahneman & Tversky, 1984; Tversky & Kahneman, 1992; Zhang & Buda, 1999). Examples abound of such negative-positive asymmetries across a wide range of domains (Baumeister, Bratslavsky, Finkenauer & Vohs, 2001; Rozin & Royzman, 2001). For example, people are more likely to report that ratios cast in the negative are more accurate (Hilbig, 2009, 2012a). However, the underlying mechanisms for explaining these differential impacts have not been extensively studied among leaders and managers, nor have the interrelationships between statistical understanding and leadership style. Indeed, some researchers have argued that a negativity bias may have adaptive value, given that negative instances are more
threatening (Labiouse, 2004; Lewicka, Czapinski & Peeters, 1992; Peeters, 2002), impelling people to react to risk in different ways.

To explain these biases, many scholars argue in favour of two different “selves” (Epstein, 2010; Kahneman, 2011; Stanovich & West, 2008) when interpreting statistics. Alter, Oppenheimer, Epley, and Eyre (2007) have posited that thought “involves two distinct processing systems: one that is quick, effortless, associative, and intuitive and another that is slow, effortful, analytical, and deliberate” (p. 569). One self is evolutionarily more primitive; the other self reflects uniquely human skills in logical reasoning, computation, and “rational” choice. The fast and frugal thinking associated with experiential and intuitive cognitive styles, which has been labelled as System 1, is considered more accessible than the laborious logic involved in the computation of numbers, associated with System 2.

These two operating systems interact in many given situations. In fact, Kahneman (2011) is explicit that System 1 (experiential-intuitive) reflects the mind’s ordinary operations. The rational analytical System 2 will intervene and critique the operations of System 1 in situations where “deliberate memory search, complex computations, comparisons, planning and choice” (p. 103) are involved. In those contexts, System 2 slows reasoning down and imposes logical analysis. Therefore in a controlled study of the interactions of frames with these two systems, the effect of the independent variable (the conditioning frame) on the dependant variable (experiential-intuitive System 1) may involve the interventions of another dependant variable (rational-analytic System 2). The effect of a frame such as reading negatively- or publicly-framed school averages may be more pronounced for individuals who are strongly experiential or intuitive in their problem-solving pre-dispositions relative to those who are more rational and analytical. Kahneman (2011) argues that System 2 is a deliberate moderator of the
caprices of System 1 in particular circumstances. The effects of framed factors on these
two systems lies at the heart of this phase of my study. These interaction effects are of
interest because they might suggest ways in which to better present test information and
better prepare school leaders. For the school principal and vice-principal, the proclivity
“to blink or think” in relation to perceived threats posed by test results becomes an
important issue in the management of meaning.

The experiment had four steps. First, 210 school administrators completed the
Rational Experiential Inventory which determined whether their problem-solving pre-
dispositions were predominately rational or experiential. Second, they supplied basic
demographic information. Third, they read one of four, randomly-assigned, student
achievement reports that were framed in alternate ways. Fourth, they completed a
scrambled version of the same Rational Experiential Inventory which re-measured
whether their problem-solving post-dispositions were rational or experiential. This
experiment was designed to determine whether school administrators’ beliefs about how
they solve problems may be altered through doubts created by a student achievement
report they read.

This experimental design thus reflects Peirce’s pragmatic ideas. According to his
thinking, the reader will consider those practical effects the average might conceivably
have. A school administrator’s conception of all these possible effects, and the
administrator’s prospective actions in terms of those effects, are thus crucial in shaping
her or his beliefs about the average. Abductive reasoning from the average as consequent
is also involved because perceptions of the average may lead to inferences about previous
conditions and actions. In an experimental design, we cannot directly document the
reader’s conception of the effects of the average. However, we can indirectly measure
the intensity of those effects by looking at changes in beliefs in relation to alternate frames for an average, as outlined in Figure 7.1 below. It is hypothesized that school administrators’ dispositions—as self-reported beliefs or habits of mind—will shift in some way because of the manner in which average student achievement is framed.

<table>
<thead>
<tr>
<th>Pre-disposition</th>
<th>Perceptual Judgment</th>
<th>Post-disposition</th>
<th>Conceived Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rational-analytic or Experiential-intuitive</td>
<td>1. Positive Confidential</td>
<td>2. Negative Confidential</td>
<td>Rational-analytic or Experiential-intuitive</td>
</tr>
<tr>
<td></td>
<td>3. Positive Public</td>
<td>4. Negative Public</td>
<td></td>
</tr>
</tbody>
</table>

**Figure 7.1.** Pragmatic model for testing effects of perceptual judgement on interpretation of average student achievement

Overall findings are described in Table 7.1 below. Across all frames, school administrators rated themselves as more rational-analytical than experiential-intuitive in their cognitive styles, both before and after reading a student achievement report. In all cases, the standard deviations within each group increased for rationality in the time interval between pre- and post- measures. In three of four groups, standard deviations increased for the experiential-intuitive scale as well. And regardless of the framed condition for presenting student achievement averages, school administrators’ mean scores declined in both rationality and experientiality after completing their reading of the
student achievement report. System 2 did not override System 1 within any of the four
groups when reading statistical information. Nor can we say that System 1 overrode
System 2 when interpreting the average in any of the four framed conditions. However,
we are dealing with very small margins: school administrators were remarkably
consistent in their self-appraisals.

Table 7.1
Pre- and Post-disposition Scores by Frame, Means and Standard Deviations

<table>
<thead>
<tr>
<th></th>
<th>Frame 1</th>
<th>Frame 2</th>
<th>Frame 3</th>
<th>Frame 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Group Size</td>
<td>51</td>
<td>52</td>
<td>55</td>
<td>52</td>
</tr>
<tr>
<td>Pre-test rational mean</td>
<td>3.94 (.38)</td>
<td>3.97 (.39)</td>
<td>3.85 (.48)</td>
<td>3.78 (.44)</td>
</tr>
<tr>
<td>Post-test rational mean</td>
<td>3.88 (.44)</td>
<td>3.96 (.41)</td>
<td>3.78 (.51)</td>
<td>3.71 (.49)</td>
</tr>
<tr>
<td>95% confidence interval for pre-test rational mean</td>
<td>3.84-4.04</td>
<td>3.86 -4.08</td>
<td>3.72-3.98</td>
<td>3.66-3.90</td>
</tr>
<tr>
<td>95% confidence interval for post-test rational mean</td>
<td>3.79-4.03</td>
<td>3.85-4.08</td>
<td>3.70-3.93</td>
<td>3.63-3.86</td>
</tr>
<tr>
<td>Pre-test experiential mean</td>
<td>3.40 (.45)</td>
<td>3.52 (.47)</td>
<td>3.33 (.38)</td>
<td>3.44 (.40)</td>
</tr>
<tr>
<td>Post-test experiential mean</td>
<td>3.34 (.50)</td>
<td>3.34 (.55)</td>
<td>3.18 (.41)</td>
<td>3.38 (.36)</td>
</tr>
<tr>
<td>95% confidence interval for pre-test experiential mean</td>
<td>3.28-3.52</td>
<td>3.39-3.65</td>
<td>3.23-3.43</td>
<td>3.33-3.55</td>
</tr>
</tbody>
</table>

*Note 1:* (Standard deviation in brackets)
*Note 2:* Frame 1 = Positive confidential; Frame 2 = Negative confidential; Frame 3 = Positive public; Frame 4 = Negative public
*Note 3:* Experiential-intuitive = System 1; Rational-analytical = System 2
7.1 Valence and Audience

To determine the impact of the four frames while also controlling for school administrators’ problem-solving pre-dispositions, I initially conducted a MANCOVA using the pre-disposition scores in a summative scale as the covariate, with separate post-disposition scales for rationality and experientiality as two dependent variables. To check whether the assumptions of MANCOVA were met, preliminary testing for normality, linearity, univariate and multivariate outliers, homogeneity of variance–covariance matrices (Box’s $M (9, 480297) =1.73, p = .08$ and multi-collinearity was conducted. No significant violation was found. Two outlier cases, which had standardized residuals greater than 3, were retained in the model because the Cook’s distance for each was less than 1, and Malahanobis distances for each were within an acceptable range of values. All MANCOVA and ANCOVA analyses were conducted using the pre-disposition summative score as covariate, presuming that rationality and experientiality are polarities on a single interval scale.

Rather than a univariate $F$ value, I used a multivariate $F$ value in MANCOVA based on a comparison of the error variance/covariance matrix and the effect variance/covariance matrix. Thus, I was dealing not with means, but rather the vectors of means within matrix algebra—these vectors would be called dynamic interpretants in Peirce’s lexicon. A family-wise Bonferroni correction was applied using SPSS facilities to preclude overly liberal interpretations. Sphericity is not a potential issue when there are only two levels (pre-post) for the measured condition. For this study, effects were considered as statistically significant at $p < .05$. Practical significance may be assumed (Cohen, 1988) when effect sizes reach .04 (.02 small, .06 moderate, 1.2 large) in a partial eta squared, as the preferred measure in educational research of this design (Richardson, 2011).
Overall effects are summarized in Table 7.2. Using Pillai’s trace, there was a statistically significant effect for time between pre-disposition and post-disposition measures, $V = .78$, $F(2, 204) = 352.89, p < .0005$, partial eta squared = .78. Approximately 78 percent of the variance in post-disposition measures was, unsurprisingly, explained by the pre-dispositional beliefs of school administrators. Likewise, there was a significant effect for public or confidential audience on school administrators’ rationality and experientiality considered together as a single dependent variable, $V = .04$, $F(2, 204) = 4.05, p < .05$, partial eta squared = .04. Approximately 4 percent of the variance in school administrators’ beliefs was explained by audience for student achievement results. Moreover, the positive-negative valence within the frames interacted with audience in a significant manner, $V = .04$, $F(2, 204) = 4.26, p < .05$, partial eta squared = .04.

Separate univariate ANCOVAs with each dependent variable revealed significant framing effects for audience on rationality, $F(1, 205) = 8.14, p < .01$, partial eta squared = .04, but not for experientiality, $F(1, 205) = 3.539, p > .05$, partial eta squared = .02. Overall, the positive-negative valence in frames for rationality was not statistically significant, $F(1, 205) = 2.45, p > .05$, nor was it statistically significant with experientiality $F(1, 205) = 54, p > .05$. Neither did the interaction effects of valence and audience have a statistically significant impact on rationality, $F(1, 205) = .79, p > .05$ or experientiality $F(1, 205) = 2.28, p > .05$. The overall univariate model explained 40% of the variance in post rationality scores (Adjusted R squared = .397), and 38% of experientiality scores (Adjusted R squared = .366). In other words, for the overall model, the audience factor was influential in inducing a small but significant overall decline in principals’ and vice-principals’ rationality. However, the overall deterioration in their experientiality was not statistically significant and was small for practical purposes.
Table 7.2

**MANCOVA/ANCOVA Summary of Effects**

<table>
<thead>
<tr>
<th>Effect</th>
<th>Pillai’s Trace</th>
<th>Omnibus F value</th>
<th>Uni F value</th>
<th>MS</th>
<th>H df</th>
<th>df</th>
<th>p value</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Multivariate Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Disposition</td>
<td>.78</td>
<td>352.89</td>
<td>2</td>
<td>&lt; .0005</td>
<td>.78</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence</td>
<td>.01</td>
<td>1.30</td>
<td>2</td>
<td>.28</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Audience</td>
<td>.04</td>
<td>4.05</td>
<td>2</td>
<td>.02*</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence X Audience</td>
<td>.04</td>
<td>4.26</td>
<td>2</td>
<td>.02*</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Univariate Effects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-disposition, Rational</td>
<td></td>
<td>127.51</td>
<td>17.10</td>
<td>1</td>
<td>&lt; .0005</td>
<td>.38</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-disposition, Experiential</td>
<td></td>
<td>114.34</td>
<td>15.40</td>
<td>1</td>
<td>&lt; .0005</td>
<td>.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence, Post-rational</td>
<td>2.45</td>
<td>.33</td>
<td>1</td>
<td>.12</td>
<td></td>
<td></td>
<td></td>
<td>.01</td>
</tr>
<tr>
<td>Valence, Post-experiential</td>
<td>.54</td>
<td>.07</td>
<td>1</td>
<td>.46</td>
<td></td>
<td></td>
<td></td>
<td>.00</td>
</tr>
<tr>
<td>Audience, Post-rational</td>
<td>8.14</td>
<td>1.09</td>
<td>1</td>
<td>.005*</td>
<td>.04</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Audience, Post-experiential</td>
<td>3.54</td>
<td>.48</td>
<td>1</td>
<td>.061</td>
<td>.02</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence X Audience, Post-rational</td>
<td>.79</td>
<td>.107</td>
<td>1</td>
<td>.374</td>
<td>.00</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Valence X Audience, Post-experiential</td>
<td>2.28</td>
<td>.306</td>
<td>1</td>
<td>.133</td>
<td>.01</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note 1.* * Statistical significance tested at $p < .05$

To further clarify the respective influences of the negativity/positivity and confidentiality/publicity factors which were manipulated in the framing, I undertook two
follow-up steps. The matrix algebra upon which are predicated such multivariate analyses of variance with a covariate implies we are testing “what if’s” as hypotheses about relationships (Garson, 2012). In other words, I may be trying to accept or reject a null hypothesis with yet another set of hypotheses. I undertook two follow-up steps because there are issues about the meaningfulness of vectors for means, and because of scholarly concerns about the appropriateness of multiple, one-way ANOVAs as the traditional follow-up method for identifying which frames or combinations of frames explain these small effects.

In the first step, I followed the advice of Jaccard & Guilamo-Ramos (2002a, 2002b) and:

1. constructed difference scores for all participants by subtracting their raw pre-score from their raw post-score (see Table 7.1) for the experiential-intuitive measure, and for the rational-analytical measure within each of the four cells as Frames.

2. ran a one-way ANOVA on these difference scores, for both the experiential and rational subscales, with the categorical variable here considered as frame group.

3. took away from these one way ANOVAs the $MS$ (mean square within groups) within each frame group or combination.

4. calculated by hand the difference scores for contrasts (parameter estimates) using the weights: +1, -1, -1 and +1 (or +.5, -.5 when combining cells), as required for the particular frames or combinations under consideration.

5. calculated the error term.

6. computed the $t$-value (parameter estimate/standard error), and generated the $p$-value, using the $t$-value and available degrees of freedom.

7. calculated the confidence interval for the parameter estimate, using the formula $PE-SE(t.crit)$ for the lower bound and $PE+SE(t.crit)$ for the upper bound.

8. computed the partial eta squared effect size for each contrast $\eta^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}}$.
These single degree of freedom contrasts are presented in Table 7.3. Overall, the table illustrates the influence of negativity as a framing factor, while also demonstrating that information in the public domain versus information confidential to teaching staff was influential, but not significantly so with a Bonferroni correction.

Table 7.3

*Single Degree of Freedom Contrasts for Rational and Experiential Subscores by Frames*

<table>
<thead>
<tr>
<th>Contrasts</th>
<th>Parameter Estimate</th>
<th>Standard Error</th>
<th>t value</th>
<th>df</th>
<th>p value</th>
<th>98% Confidence Interval</th>
<th>$\eta_p^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frame 1+3 versus Frame 2+4 for rational</td>
<td>-.155</td>
<td>.064</td>
<td>2.42</td>
<td>106</td>
<td>.008*</td>
<td>-.219, -121</td>
<td>.05</td>
</tr>
<tr>
<td>Frame 1+2 versus Frame 3+4 for rational</td>
<td>-.06</td>
<td>.064</td>
<td>.93</td>
<td>104</td>
<td>.177</td>
<td>-.124, -.004</td>
<td>.01</td>
</tr>
<tr>
<td>Frame 1+3 versus Frame 2+4 for experiential</td>
<td>-.104</td>
<td>.062</td>
<td>1.67</td>
<td>106</td>
<td>.048</td>
<td>-.166, -.042</td>
<td>.03</td>
</tr>
<tr>
<td>Frame 1+2 versus Frame 3+4 for experiential</td>
<td>.0</td>
<td>.062</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
</tbody>
</table>

*Note 1.* Frame 1 = Positive confidential; Frame 2 = Negative confidential; Frame 3 = Positive public; Frame 4 = Negative public.

*Note 2.* Experiential-intuitive = System 1; Rational-analytical = System 2.

*Note 3.* In each contrast, the Post score is subtracted from the Pre score and divided by two to yield a parameter estimate.

*Note 4.* * Critical $p$ threshold adjusted to .0125 under the Bonferroni procedure to control for family-wise error rate within four contrasts.

Contrasts between various combinations of reading tasks show that frames 1 and 3 together were significantly different from frames 2 and 4 together in reducing school administrators’ rationality in problem-solving dispositions, $t(1,106) = 2.42$, $p < .0125$, partial eta squared = .05. Those readings that framed results in the negative had a significantly larger impact on school administrators’ dispositions, reducing an inclination to be logical-analytic, but not an inclination to be experiential. With a .0125 threshold for significance, no significant differences were apparent for frames 1 and 2 in combination versus frames 3 and 4 in combination—in other words, it did not make a
difference whether the test information was made public or held confidential to the school in changing school administrators’ cognitive styles.

As a second step, to isolate the particular frames having an impact, and to corroborate the findings from single degree of freedom contrasts, I conducted an ANCOVA using the post-disposition measure for rationality as a single dependent variable, with the summative pre-disposition measure as covariate. Then I conducted a separate ANCOVA using the post-disposition measure for experientiality, with the same summative pre-disposition covariate. Bonferroni adjustments were made within SPSS facilities. Table 7.4 summarizes simple contrasts from this analysis. For rationality, the hypothesized difference scores between Frame 2 (negative confidential) versus Frame 4 (negative public) were statistically significant $H(2, 103) = .186, p < .05$. A negative student achievement report that was confidential to the school had a greater impact in reducing principals’ and vice-principals’ rationality than did a negative report that was anticipated to become public. All other contrasts for both rationality and experientiality were statistically non-significant.

Table 7.4

**ANCOVA Contrast Results (K Matrices)**

<table>
<thead>
<tr>
<th>Simple Contrast</th>
<th>Estimated Difference $H$</th>
<th>$SE$</th>
<th>$p$ value</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Rationality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frame 1 vs Frame 4</td>
<td>.065</td>
<td>.073</td>
<td>.38</td>
<td>-.079, .208</td>
</tr>
<tr>
<td>Frame 2 vs Frame 4</td>
<td>.186</td>
<td>.072</td>
<td>.01*</td>
<td>.045, .328</td>
</tr>
<tr>
<td>Frame 3 vs Frame 4</td>
<td>-.036</td>
<td>.072</td>
<td>.62</td>
<td>-.177, .106</td>
</tr>
<tr>
<td><strong>Experientiality</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Frame 1 vs Frame 4</td>
<td>.058</td>
<td>.073</td>
<td>.42</td>
<td>-.086, .202</td>
</tr>
<tr>
<td>Frame 2 vs Frame 4</td>
<td>.007</td>
<td>.072</td>
<td>.92</td>
<td>-.135, .149</td>
</tr>
<tr>
<td>Frame 3 vs Frame 4</td>
<td>-.105</td>
<td>.072</td>
<td>.15</td>
<td>-.247, .037</td>
</tr>
</tbody>
</table>

Note 1. All hypothesized differences control for pre-disposition as covariate

Note 2. Statistical significance tested at $p < .05$
Drawing together these three different analyses, then, we can observe that:

1. Pre-dispositional beliefs prevail overall in school administrators’ interpretations of the average; 2. The positive-negative valence of test results alone does not significantly change administrators’ beliefs, but does so significantly in interaction with the anticipated audience for that information; 3. Rationality deteriorates significantly under all framed conditions; 4. Experientiality deteriorates but not significantly so under most framed conditions, and; 5. Significant interactions between valence and audience (i.e., negative, confidential reports) clearly erode the rationality and more ambiguously the experientiality of Saskatchewan school administrators.

Thus, we can say that how reports are framed has a small and immediate effect on Saskatchewan school administrators’ problem-solving dispositions, leading them to adopt less logical-rational approaches in all framed conditions. Experientiality likewise deteriorates, but the statistical significance of that drop depends on the threshold accepted. Interactions between frame and rationality were evident, but not in the direction forecast by Tversky and Kahneman (1992) who claimed that the reading and perhaps computation of numeric information for planning and comparative purposes causes adults to slow down and become more logical and analytical, derogating from more intuitive and heuristic approaches.

In short, negatively-framed information about average student achievement before a within-school audience has salience for school administrators’ problem-solving styles, reducing rationality. In contrast, the prospect of such information entering the public domain rather than remaining confidential to the school did not appear particularly influential in shaping school administrators’ approaches to problem-solving and planning.
Nevertheless, any effects appear small. An overall pattern of continuity was apparent across each group in time, although expanding standard deviations suggest that effects may be more pronounced at the individual level than at the group level.

Therefore, the hypothesis (2.a) that school administrators will adopt more rational and analytical problem-solving dispositions when school averages are framed negatively than when they are framed positively is not accepted. In fact, the trends here suggest that school administrators would adopt less rational and analytical problem-solving dispositions when school performance is cast in a negative light, while sustaining experiential or intuitive dispositions. Likewise, the hypothesis (2.b) that school administrators will adopt more experiential and intuitive approaches to problem solving when school averages remain confidential to the school rather than being made public is not accepted. Only a slight and overall minimal change was evident in school administrators’ tendency to address planning and problem-solving with intuitive-experiential approaches when school averages were expected to become public.

Nevertheless, there were significant and complex interactions between valence and audience as student achievement averages were interpreted by school administrators.

7.2 Occupational Position and Disposition

To determine whether vice-principals hold more rational and analytical pre-dispositions to the interpretation of averages than do principals, I conducted a one-way ANOVA with the summative pre-disposition scale as dependant variable, using administrative position as an independent categorical variable. Levine’s test of homogeneity of variance was computed and found to be non-significant. The difference between the means of principals and vice-principals was not significant,
$F(2, 207) = .941, \ p < .39$. Vice-principals did not significantly differ from principals in their pre-dispositions to plan for improvements in student achievement with anticipated results in hand. Therefore, the hypothesis (2.c) that vice-principals will have more rational and analytical pre-dispositions to the interpretation of average school performance than will principals is not accepted. In short, administrative position does not relate to any proclivity to be either rational or experiential in one’s problem-solving within a school setting when addressing a student achievement issue.

7.3 Demographics and a Descriptive Model

To determine which demographic variables best predicted school administrators’ pre-dispositions when planning for the solution of a student achievement problem, I performed a multiple regression where the six demographic variables were entered simultaneously into the model, this time using the summary pre-disposition scale as dependant variable. The total sample size of 210 administrators should have the statistical power to detect medium to small effect sizes (Miles & Shevlin, 2001). Preliminary analyses were conducted to ensure no violation of the assumptions of normality, linearity, multicollinearity, and homoscedasticity.

This multiple regression analysis is summarized in Table 7.5 below. Taken together, the six variables predicted only seven percent of the variance, $F(6, 201) = 3.67, \ p < .01$ in school administrators’ pre-dispositions to be rational-analytical or experiential-intuitive in their approach to problem solving, when considered on a single continuum. Among the array of variables entering into the model, only statistical self-efficacy made a unique and significant contribution (beta = .09, \( p < .001 \)) to the model.
Table 7.5

Multiple Regression of Demographic Variables with Saskatchewan School
Administrators’ Problem-Solving Pre-dispositions

<table>
<thead>
<tr>
<th></th>
<th>B</th>
<th>SE B</th>
<th>β</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>2.98</td>
<td>.19</td>
<td></td>
</tr>
<tr>
<td>Gender</td>
<td>-.08</td>
<td>.04</td>
<td>-.13</td>
</tr>
<tr>
<td>Position</td>
<td>-.02</td>
<td>.04</td>
<td>-.03</td>
</tr>
<tr>
<td>Admin experience</td>
<td>-.04</td>
<td>.02</td>
<td>-.15</td>
</tr>
<tr>
<td>Educator experience</td>
<td>.05</td>
<td>.03</td>
<td>.16</td>
</tr>
<tr>
<td>Post-sec education</td>
<td>.01</td>
<td>.02</td>
<td>.04</td>
</tr>
<tr>
<td>Statistical self-efficacy</td>
<td>.09</td>
<td>.03</td>
<td>.24*</td>
</tr>
</tbody>
</table>

Adjusted $R^2$ for total model = .07 ($p < .01$). *$p < .001$

Note. Educator experience = teaching and administrative experience combined.

An examination of $B$ and $\beta$ values yields further insights. Of particular interest are the negative values for gender, position and administrative experience, which indicate these background characteristics had an inverse, predictive relationship with the overall administrative pre-disposition to be predominately rational. That is, when the respondent was female, or when the respondent was seated in a principal’s chair rather than a vice-principal’s, or when the respondent had acquired more administrative experience, the less the characteristic was related to a rational pre-disposition and the more it was related to an experiential pre-disposition. However, overall educator experience—which combines teaching and administrative experience—made a positive contribution to the model, if we presume that greater rationality is the desired outcome. In other words, accumulated administrative experience and total accumulated professional experience (teaching plus administrative experience) appeared to operate in opposite ways on problem-solving pre-dispositions. Administrative experience contributed to more intuitive and experiential
problem-solving approaches. In contrast, more combined teaching and administrative
experience led to more logical and analytical approaches.

Because standardized beta values are measured in standard deviation units and
thus are directly comparable, they provide better indications of the relative importance of
the predictors. The most important predictor was statistical self-efficacy. Self-appraised
statistical knowledge was significantly and unsurprisingly related to more rational and
analytical problem-solving styles. Accumulated administrative and professional
experience followed as being equally important, these latter two working in opposite
ways on dispositions as noted above. And finally, the gender of the administrator was
relatively important compared to position and post-secondary education levels, but not
significantly so in this predictive model. Consistent with a previous study (Peters &
Levin, 2002), in this study, education level and numeracy did not predict styles. Because
the residuals in this model were large, we can say that other factors—such as personality,
school size, school type (elementary or secondary), geographic location, or district office
policy and leadership expectations—may be much more important in shaping the
administrators’ mindset than those variables identified in this study.

Therefore, the hypothesis (2.d) is modified to read that experience as a teacher
and administrator, position (principal or vice-principal), gender, education level in
themselves do not substantially or significantly predict school administrators’ problem-
solving pre-dispositions when reading statistical averages. However, self-efficacy
(considered as the administrators’ self-appraisal of their own statistical knowledge)
makes a unique and significant contribution to a model for predicting school
administrators’ pre-dispositions when reading averages in order to problem-solve and
plan around student achievement.
7.4 Discussion and Implications

In the psychophysics of perception, Peirce recognized that conditions in the environment will alter people’s perceptions of phenomena. School administrators’ conceptions of the average, and hence their dispositions in problem solving appear to be subject to some of those same contingent influences. Nevertheless, the most discernible pattern in this study was the consistency between school administrators’ dispositions after reading and reasoning with averages and their pre-dispositions before undertaking the experimental task. Their pre-reading beliefs about their self-reported managerial approaches explained nearly four-fifths of the variance in their post-reading beliefs, whereas the student achievement report itself explained only five percent. Regardless of whether an anticipated normative distribution of student achievement results focuses on students who are below or at average or on students who are above average, on results that are in the public domain or confidential to the school, only small effects were detected after reading the average within one of the four alternate frames. Witteman, van den Bercken, Claes, and Godoy (2009) have also noted significant interactions between the rational and experiential dimensions of the mind in small, non-orthogonal studies of European students given computational tasks, with the rational becoming dominant.

Peirce accepted that Fechner had established a “law of the mind” in which the intensity of any sensation is proportional to the logarithm of the external force which produces it. However, Peirce defined the law such that “as the physical force of excitation of a nerve increases geometrically, the sensation increases arithmetically, so that the sensation is proportional to the logarithm of the excitation.” Fechner, in contrast, argued that the total sensation varies directly with the logarithm of the stimulus divided by the stimulus just sufficient to give an appreciable sensation (EP1.159, 379). Peirce was not consistent in all his writings that this zone underneath the threshold for “just sufficient” was the pre-conscious seat for abductive insight and hence where belief begins. For Peirce’s (and Jastrow’s) experimental refutation of Fechner’s law, see W5.122-135. Kahneman and Tversky’s Prospect theory (1984) replicates Fechner’s findings in the realm of behavioural economics. Although the current experimental design did not permit testing either “law” for degrees of belief in problem-solving dispositions with statistics, one might manipulate the intensity of various frames by incrementally (arithmetically or geometrically) adjusting the magnitude of “excitation” found in either publicity/confidentiality or positivity/negativity.
However, this finding in Witteman et al.’s study was not replicated here. It is important to note that all four data displays in this study were statistically, graphically, and formally identical in their content, and thus it should not be surprising that effect sizes will be small.

Nevertheless and notwithstanding the small effect sizes, it appears that school administrators are susceptible in some measure to framing effects like other adults. Even with extensive post-secondary education and decades of experience in dealing with student marks, many school administrators demonstrated that they will shift slightly in their cognitive styles when faced with the prospect of negatively-framed student achievement. Two alternative explanations for this proclivity to over-weigh negative information (Hilbig, 2009) have been identified. An elaboration hypothesis suggests that negative information stimulates respondents into more effortful processing and constructions which render the statistic more compelling. The alternative hypothesis is “information processing fluency”—that is, negative information is processed faster than positive information, since it attracts immediate attention. These two alternative mechanisms, (1) elaboration and (2) fluency, thus make distinct and even contradictory claims about the functioning of an individual mind which has been characterized as predominately System 2 or System 1 in situations where statistics are interpreted. The evidence in the current study lends credence to the fluency hypothesis. When focusing on the bottom end of a distribution of student test scores while interpreting results, the school administrators’ proclivity for abstract, analytical thinking was reduced, while initial feelings about people, trust in snap judgements, or gut feelings about people changed only in non-significant ways.

Although only the changes in administrators’ rational-analytical dispositions were statistically significant, alternate but not mutually exclusive explanations may be offered
for the overall decline in both rational-analytical and intuitive-experiential approaches following a reading of the average. Three explanations accord with Peirce’s pragmatism. The first, regression to the mean, relates to experimental design (Nesselroade, Stigler, & Baltes, 1980). Regression toward the mean simply says that, following an extreme random event, the next random event is likely to be less extreme. Researchers since Galton’s time have noticed this measurement phenomenon. Both Peirce elder and junior promoted the method of least-squared means for addressing the issue. What is known as Peirce’s criterion for eliminating outliers from analyses to correct for this measurement phenomenon touched off an international debate about the practice in the 1870s (Peirce, 1870; Rovine & Anderson, 2004, 2011). The converse may also hold: events that register as average on the first measure may egress and be registered as more extreme on the second measure. In fact, the expanding standard deviations in post-disposition measurements for all four frames contradict regression to the mean as an explanation. Moreover, measurement error was considered in each analysis of significance.

A second explanation for the eroding means may be that school administrators’ readings of performance data induce neither an exclusively rational-analytical nor intuitive-experiential problem-solving style, contrary to current propositions in cognitive psychology. Rather, reading statistics invokes an interactive amalgam of both, or perhaps a third and independent mode of inferential reasoning altogether. This hypothesis is supported by the concurrent declines in scores for both System 1 and System 2 “selves”. There may be other logics at work, not just the deductive modes of mathematics or inductive modes of intuitive reasoning. That is, school administrators may not consider abductive reasoning as a form of logic or as a form of experience at all. However, abductive reasoning was not measured in this study, and indeed may not be measurable.
A third explanation is that declining means under all four frames reflect the impact of doubt. What the reading task has generated is not administrative conviction, nor changed beliefs about what constitutes average student achievement, but rather uncertainty about school administrators’ plans, about the veracity of the predicted outcomes, or about the appropriateness of their projected course of action. Given the significant interactions between valence and audience as framed evidence with the reasoning processes involved in System 1 and 2, we might conclude this is a genuine and reasonable form of doubt, not the feigned doubt that Peirce discussed as inherent within some forms of scientific inquiry (EP1.127-128; EP2. 338).

A fourth explanation may also be offered, that does not accord with Peirce’s ideas. Although rationality declined under all four framed conditions, it declined less so than experientiality under the negative and confidential frame, suggesting that school administrators were reluctant to dispense with a self or projected image of rationality when completing an inventory of their beliefs. The experiment depends on self-reported beliefs; image management may account for the results. However, Peirce argued that we should not believe witnesses “are falsifying, without any definite, objective and strong reason for the suspicion” (EP2. 113). We must prima facie accept principals’ responses on the Rational Experiential Inventory as genuine.

The student achievement report itself was not normatively symmetrical in this study, even if the four frames were graphically and statistically orthogonal: 80 percent below and at average is mathematically identical to 20 percent above average in an asymmetric distribution. The report itself did not frame statements of probability or frequency, but rather statements of percentages and proportions with criterion statements of average student performance. If so, a “ratio-bias” is evident not only in adults’ estimates of probability but also when they perceive the mean within a skewed
distribution (Reyna & Brainerd, 2008). Stated another way, perceptual judgements about “a school glass” being 4/5 empty overweigh an appraisal of the glass being 1/5 full, obscuring the 2/5 constant in the middle within an identical distribution, leading to changes in administrators’ dispositions. These biases may also explain, in part, the related and important difficulty that well-educated adults in managerial positions have in gauging accumulation, as stock flows within an organization—whether it is the movement of merchandise through a factory or students through a multi-year program of studies (Cronin, Gonzalez, & Sterman, 2009). The ability to plan and organize inputs, within inputs and outputs may be impaired, particularly if managers do not hold a flow conception for the average. In an education context, understanding the accumulation of credits and courses is particularly crucial when making decisions about the movement of students through various curriculum objectives and through the various grade levels. Such ideas about the average become vital when school administrators are making decisions about student promotion and retention.

Several alternate theories have been formulated to explain this bias—class inclusion difficulties (confusion associated with overlapping categories when perceiving objects), the predominance of qualitative gist over verbatim statistics in the individual memory (retrieving essentialist rather than detailed images in working memory), and denominator neglect when performing arithmetic calculations to weigh risk (overweighing the importance of the numerator). This experiment corroborates the first theory—class inclusion difficulties. School administrators’ dispositions were affected more when below average students were grouped with average students, than when the administrators were invited to focus on above average students alone. However, we cannot attribute this bias to either a rational-analytical or experiential-intuitive disposition, but only to the interactions between the two.
In classical pragmatism, class inclusion difficulties stem from the mixture of purposes that any interpreter brings to bear when reading a statistic. Problems in categorization were recognized by Peirce as complications in “the general plan of desire.” Desires create our categories, he argued. However, desires are vague. Because our desires or purposes are always multiple, we must strike compromises among them and so our classifications start to overlap (EP2. 115-119). What is discerned as average student achievement thereby becomes blurred with achievement considered to be below average. In this experimental design, however, we are only indirectly able to identify such confusions arising from a reader’s contradictory purposes when interpreting an achievement report.

Although contemporary cognitive-experiential theorists (Epstein, 2010; Pacini & Epstein, 1999) emphasize the experiential (intuitive) origins of such biases, many predictions involve both rational and experiential processes. Distortions in numerical processing, such as frequency and ratio biases, “can result from strong experiential processing, weak rational processing, or both” (Pacini & Epstein, 1999, p. 981). In this line of argument, ratio concepts are especially difficult when weighing risks because experience and intuition interfere with “correct” processing or because rational processing (e.g., numerical computation) is inadequate. These predictions are qualified, however, for “non-optimal” ratio tasks. According to Pacini and Epstein (1999), “The safest prediction that can be made when optimality is at issue is that non-optimal responses will be inversely related to rationality, directly related to experientiality, or both” (p. 981). Thus the interaction effects observed here, which suggest declines in rationality and overall stability in experientiality, only partially support cognitive theories of personality (Epstein, 2010). Numeric reasoning derives from various mixtures of both systems, depending on the individual, underlining Peirce’s synechistic ideas about underlying continua.
Some researchers have argued that accountability for outcomes has more pronounced impacts on school administrators’ judgement than does accountability for processes, where the administrator is responsible for only the planning mechanisms within a school (Langhe, van Osselaer, & Wierenga, 2011). In this experiment, the school administrator as subject was provisionally but explicitly assigned responsibility for planning processes, not for the achievement of students. Nevertheless, it remains difficult to distinguish between responsibility for processes and for outcomes. Even when given responsibility only for planning and not for results, school administrators will demonstrate small changes in leadership style, depending on the valency for what is considered as average. When a school administrator’s job depends on improved student performance, the impact of a particular statistic on judgement and problem-solving ability may be increased.

Although many questions have been raised about distinct leadership styles by gender in school administration (Grogan & Shakeshaft, 2011), few clear answers were provided in this study. Gender differences in problem-solving and decision-making styles were small, and statistically non-significant in predicting school administrators’ problem-solving pre-dispositions. Only minor differences were evident, although female administrators reported a slightly greater preference for an experiential approach (Sinclair & Ashkanasy, 2005).17 Given the bounds of this experiment, no conclusions may be drawn about gendered approaches to university programs in mathematics, statistics, or leadership preparation.

17 In Peirce and Jastrow’s (1885) experiment, they probabilistically found that subjects had sensations below the threshold of conscious perception. That fact had “highly important practical bearings”, they believed, since it gave “new reason for believing that we gather what is passing in one another’s minds in large measure from sensations so faint that we are not fairly aware of having them, and can give no account of how we reach our conclusions about such matters.” They speculated that “the insight of females as well as certain ‘telepathic’ phenomena may be explained in this way” (p. 83).
However, many implications follow from the finding that school administrators were more likely to change their problem-solving dispositions after reading negative test results which are kept confidential to teaching staff, than when a set of test outcomes are positive and made public, becoming less rational and analytical in their approaches. Principals and vice-principals did not report becoming either more logical-linear or intuitive-experiential when test results were anticipated to be negative and to become public. Negative images for student achievement averages appeared to call for less abstract, less logical, and more flexible approaches to instructional leadership with teachers than with parents or the general public. In other words, the challenges for managing the meaning of below average student performance appear to reside primarily within the school rather than in the larger community. Feeling, gut instinct, and the heart remained important for guiding actions and planning with test results within the school.

Apparently contradictory findings about the influence of particular frames may be reconciled by considering the assumptions that undergird particular data analysis techniques. MANCOVA and ANCOVA operate at the macro-experimental level, dealing with vectors of means, computing interactions therein at the individual respondent level. Single degree of freedom contrasts among combinations of groups presume that raw mean scores are subtractive across framed conditions, not point estimates on a slope as in MANCOVA. Straightforward comparison of means within a particular framed group (particularly under the negative confidential frame) assumes we may compare intact group means between rationality and experientiality within a solitary set of conditions—ignoring correlations, covariance, standard deviations and thus variation. Rather than a point estimate as in MANCOVA, the mean becomes a statement of typicality. Each technique puts different boundaries on bounded rationality. The meaning of a mean depends on method, and a parallax in interpretation may arise.
Moreover, as Peirce insisted, terminology is important when defining conduct (EP2. 262-266). It makes a substantive difference in both research design and in practice whether we characterize the rational-analytic and the intuitive-experiential systems as operating independently (distinct systems), autonomously (parallel and in tandem in ongoing consideration of the other), interactively (occasionally interrupting and redirecting each other), or interfering (disrupting or impeding) across a pair of continua. Likewise within a single continuum, the rational-analytic and intuitive-experiential dimensions of the psyche can be conceived as polarities (at the expense of each other) or as synthetic (as an amalgam). A key issue is whether we are dealing with interplay or cross-play. Thus, more precise information is required about whether rationality intervenes, interrupts, over-rides, moderates, tempers, supplants, encompasses, or simply verifies the output of the experiential dimensions in ordinary thinking, and under which circumstances. Each in turn has implications for pedagogy, ranging from how adult educators sequence instruction through to how they approach remediation in mathematical understanding. For school leaders who are “controlling” the meaning of student marks, the distinctions may imply very different strategies for engaging others, ranging from the dialogical to the dialectical. “There is no distinction of meaning so fine as to consist in anything but a possible difference of practice,” Peirce argued (EP1. 131).

Any study of effect sizes also involves judgements about the importance of an effect for leadership practice (Grissom & Kim, 2005). Even small effects may be important in experiments where difficult-to-change constructs such as personality are involved, or where minimal manipulations of an independent variable are undertaken (Prentice & Miller, 1992). Both apply here. Leaders’ dispositions to be rational or experiential may work across a variety of domains, and can be resilient to change. In this study with contrived conditions, framing the task involved switching only the valency
(positive/negative) for an average and the anticipated audience for school outcomes (public or confidential). Statistically significant effects on problem-solving styles followed, but the effect sizes were small. These results have practical implications for the conduct of school administrators, suggesting that external controls are required in data-informed decision-making, and for graduate educational administration programs, where matters of gut feeling, insight and instinct should be treated seriously in conjunction with statistics. The multiple regression model might support those who claim that more attention be given to statistical self-efficacy in undergraduate teacher preparation and graduate educational leadership programs (Finney & Schraw, 2003). This does not necessarily presume the logical-analytical approach is superior to an intuitive-experiential approach when interpreting statistics. Nevertheless, it does argue for more pragmatically-oriented programs which align with both the authors’ and readers’ purposes, not just upholding the statistician-authors’ intentions and interests.

Before conclusions are drawn, it is necessary to recognize this experiment’s limitations. First, the reading task asked principals to imagine a scenario wherein their school receives results that are asymmetrically distributed, not one that describes actual results received by the school, which may be skewed in different ways. When accomplishing the reading task, administrators were working in the conditional future tense (If I would have these results), not in the present tense (I do have these results) with test outcomes that were actually registered for their school. Second, school administrators were working with information that was presented simultaneously in graphic, textual and numeric formats. Thus, it is not possible to determine whether

18 Peirce and Jastrow (1885) contended that sensations just below the level of awareness “ought to be fully studied by the psychologist and assiduously cultivated by every man.”
school administrators were responding to the average as histogram, as prose, or as numeric figure. Third, the task invited administrators to read and interpret averages without requiring them to compute or do actual calculations with pencil and paper or calculator. Stronger effects may have been evident if, in one or more frames, school administrators had been required to directly compute proportions of the student population reaching particular levels, thereby more directly calling on the rational System 2 to override or derogate from the heuristic judgements of System 1. Fourth, none of the frames specify a particular grade level or subject area or source for the test results—whether the test results came from within the school or from an external source such as a school division-wide or provincially-mandated test. Clarifying the source of the information may either accentuate or attenuate the impacts described here. Fifth, the task did not provide precise information about what is involved in making the information public or keeping it confidential. The label “public” can defined many ways, including community gossip (McColskey, Altschuld & Lawton, 1985), publicity in the local or provincial mass media, think tanks’ re-amplifications, or distributing test scores in a controlled fashion to a more narrowly-defined public body such as a school community council or school board. Likewise, there may be varying degrees of confidentiality within a teaching staff—ranging from a brief mention at a staff meeting, to sharing summary information among school departments via memo, to sharing detailed test results with all school professional and support staff at an inservice session. The role of the school administrator is, by its nature, a public role. Being more specific about the type and degree of both publicity and confidentiality given to a student achievement report becomes an important variable for future studies. And sixth, we must recognize
that only family-wise and not experiment-wise error rates can be controlled in research of this design.

Even with these constraints, several points can be provisionally made. School administrators’ perception and hence conception of the average appears to depend on the way that average student achievement or the middle is framed. The main message is that even moderate changes in administrators’ cognitive styles when addressing student achievement issues would not automatically follow their reading of a single achievement report. Nor would school administrators change markedly in their problem-solving pre-dispositions when school results are about to become public rather than staying confidential to the school. However, with a set of results which are perceived as below/at average, school administrators’ reliance on careful logical analysis may be significantly decreased in the immediate term and, seemingly, with little impact on their proclivity to rely on gut feeling, instinct, and the heart as guide. Negatively-framed student test results may have small impacts, but administrators are unlikely to over-engage in laborious logical-analytical processes of calculation. Nor would they become overly rigid in their problem-solving within the school. If anything, negative school results may elicit less, not more reliance on “logic” and “math” as a response. Expanding standard deviations between pre- and post-reading dispositions indicate that, across Saskatchewan, diversity in administrators’ responses would be retained and even amplified. We must recognize that the absence of evidence is not the same as evidence of absence. Nevertheless, there is no evidence in this study, at least in the short run, that reading and interpreting averages framed as negative in the public domain would induce the more analytical, rigid and inflexible modes of thinking implicit in deductive reasoning.
In a 1903 Harvard lecture entitled “Pragmatism as the Logic of Abduction,” Peirce noted that, “the contents of the perceptual judgment cannot be sensibly controlled now, nor is there any rational hope that it [sic] ever can be” (EP2. 191; 240-241). In this pre-conscious, instinctive part of the mind, logical maxims have about as much applicability as they do to the growth of our hair and fingernails, he claimed. We “as nicely adjusted organisms whose equilibrium has narrow ranges of stability” (EP2. 241) are susceptible to a variety of conditional frames, including accidental circumstances in the moment. Recently and more optimistically, Kahneman (2011) has argued that the caprices of experiential and intuitive impulses when reading statistics may be moderated in management situations by fellow professionals, substituting peer controls for the loss of self-control which is characteristic of snap judgment. In other words, group decision-making is recommended when there are negative statistics. For those wishing to manage the meaning of a negative statistic in a reasoned manner, this recommendation appears both sensible and pragmatic.

7.5 Recapitulation

In light of this study’s conditions and limitations, we might surmise that:

1. Perceptual judgement is related to school administrators’ conception of an average as depicted in a numeric, graphic and prose representation.

2. Saskatchewan school administrators’ problem-solving styles would generally remain constant when they are given the task of planning for improved average school achievement, and asked to interpret a statistical average, regardless of how it is framed.

3. Perceptions of below/at average student achievement at the school level would have small effects on school administrators’ problem-solving dispositions. Negatively-framed information would reduce rationality as defined within a formal, logical type of reasoning, without probabilistically diminishing a propensity to be experiential or intuitive.
4. The prospect of school-level student achievement averages entering the public domain rather than remaining confidential to teaching staff would not particularly or immediately influence school administrators’ problem-solving and planning approaches.

5. Those student achievement averages which are cast in a negative light and are confidential to the school may induce significantly less rational-analytical dispositions in school administrators than averages which are cast in a negative light and made public.

6. Vice-principals would demonstrate similar dispositions to principals in their problem-solving in relation to average student achievement.

7. Although a school administrator’s gender, position, accumulated administrative and teaching experience, and level of post-secondary education would influence in a small way the administrator’s pre-disposition to be rational or experiential, only statistical self-efficacy would strongly predict an administrator’s proclivity to be more logical-analytical in his or her problem-solving.

8. Peirce’s propositions about perceptual judgement and its effects on conception are corroborated within his theory of interpretation. However, many of the influences on school administrators’ beliefs and problem-solving approaches with statistics of various types have yet to be identified.
Chapter 8: The Significance of Average Student Achievement in School Administrative Practice

This chapter draws together the findings of the mixed-method study designed to help create a well-rounded view of the effects of averages for instructional leadership. Peirce’s criteria for appraising both practical and statistical significance are described, and conclusions are drawn about the applicability of classical pragmatism for future research into statistical interpretation.

8.1 Synopsis of the Study’s Purposes and Findings

Understanding school administrators’ conceptions, interpretations, and actions relating to elementary statistics is important for several reasons. First, school principals are both professional and community leaders who model numeracy, a widely recognized goal in most public education systems. Second, Saskatchewan and Canadian adults’ ability to access, use, interpret and communicate mathematical information for managing or solving a real-life problem does not meet international norms (Statistics Canada, 2013; OECD, 2013). Third, a major presupposition of classroom assessment and large-scale assessment programs across North America is that some form of educational good follows from principals’ interpretation of educational statistics. Fourth, understanding how school principals conceive of a statistical average helps us understand how these averages affect activity within schools. Fifth, principals’ ability to see themselves as agents of change and to use performance statistics as guideposts for change is increasingly recognized in leadership studies as key for effective change. Sixth, preparing school principals to move from a managerial to a leadership outlook implies that we need to understand their current perspectives in order to develop effective
educational leadership programs. Seventh, scholars have recognized, but not extensively investigated, the different colloquial or quotidian meanings held by adult readers for many statistical ideas communicated in government reports and in the media (e.g., random, representative, probable, reliable, confidence), nor have scholars investigated the way readers respond to these ideas (Gal, 2002).

Knowing whether school administrators become predominately rational-analytic or experiential-intuitive when interpreting statistics has many implications for cultivating the statistically-knowledgeable school leader (Eberly & Fong, 2013) (see Appendix J). If we are to improve student outcomes in Canadian schools, some knowledge of what and how school principals think about average student achievement is pertinent (Bush & Middlewood, 2005). Measures of central tendency appear in many forms of educational research, budget statements, and most assessment findings. More precise notions of the average student should create a clearer picture of the exceptional student and thus enable greater focus in allocating resources.

To explore school administrators’ conceptions of average student achievement, I first interviewed 10 Saskatchewan school principals in diverse school settings and invited them to read aloud three alternate data displays depicting averages. In a succeeding experiment, 210 school administrators completed a self-inventory of their problem-solving dispositions as beliefs before and after reading an average framed in four different ways. Both phases of this study were informed by Charles Sanders Peirce’s pragmatic theory of interpretation.

Overall, most school principals considered student averages as the starting point for further interrogation or as a platform for discussion, rather than as a conclusive depiction of actual qualities or an unimpeachable statement of truth. In relation to the
two research questions presented in Chapter 2 with initial conjectures and the formal hypotheses, I found that:

1. **What are school administrators’ conceptions for the term “average” student achievement?**
   
a. School principals held nuanced conceptions of the average in multiple representations of average student achievement, but did not draw on mathematically-sophisticated concepts for either measures of central tendency or a norm when interpreting student achievement reports.

   b. Principals conceived of the average as a centre-of-balance, but understood the computational average as an arithmetic mean weighted in relation to curriculum priorities for individual students, as a midrange or modal distribution for grade level or classroom marks, and as a simple arithmetic mean of several key curriculum domains for school level results, even when results were presented in alternative formats with alternative statistical forms.

   c. Principals were aware of the long-term socio-educational effects of report card averages, and therefore put a priority on classroom marks rather than on large-scale assessment results. Their reasoning was largely abductive for any type of assessment result, moving from the average as consequent to considering anterior conditions.

   d. Principals did not discount notions of significance as the comparison of two means, but most often considered averages across a variety of assessments rather than within a single data set. Those averages deriving from budgetary allocations and class sizes such as PTR were considered more relevant when defining instructional arrangements than student outcomes.

2. **How does framing “average” student achievement, in the positive and as a statistic made public, influence school administrators’ conceptions of the average and their problem-solving approaches?**

   a. School administrators adopted less rational and analytic problem-solving dispositions when school averages were framed negatively than when they were framed positively.

   b. School administrators did not adopt either more rational-analytic or more experiential-intuitive problem-solving approaches when school averages were framed as confidential to the school, than when they were framed as in the public domain.

   c. Vice-principals’ cognitive styles were similar to principals’ when interpreting average school achievement.

   d. Experience as a teacher and administrator, position, gender, and education level did not uniquely predict school administrators’ problem-solving dispositions when reading statistical averages. However, statistical self-efficacy did.
8.2 Conceptions and Perspectives on the Average

At the outset of his fertile career, Peirce acknowledged that thought can be considered from two vantage points: as inferential content or as psychological activity (EP1. 8; EP1. 174). It follows, then, that the significance of average student achievement for school administrators may be ascertained in terms of a conception’s substantive features, or in terms of the psychological processes engaged when interpreting a statistic. In Peirce’s theory of interpretation, concepts are colligations of signs bound together to furnish an explanation. Conception involves receiving, producing or reproducing thought with such Signs as representations in practice. The interpretation of Signs, whether linguistic or arithmetic or graphic, involves not only intellectual but also somatic and emotional responses. In a Prospect theoretic account within cognitive psychology, on the other hand, statistical interpretation involves understanding the operations of the rational-analytic and experiential-intuitive selves in decision situations. Each theory offers alternative but complementary vantage points for examining school principals’ management of the meaning for average student achievement within instructional leadership.

According to Peirce, a single Sign is amenable to multiple conceptions depending on whether it is construed initially as qualitative icon, as an indexical point of comparison, or as a symbolic notation. Peirce’s pragmatism rests on a straightforward premise: a conception reduces the manifold of sense impressions to unity (EP1. 1; CP1. 545; W2. 49). The manifold is a mathematical or topographical idea for a collection of points as Signs in space. However, Peirce argued that it is impossible to render such impressions to unity in a single conception in the consciousness. In becoming conscious of the initial conception alongside the manifold, another conception is created, and so on. Therefore, three primal gradations of clarity or categories of conceptions arise—quality,
relation, and representation (EP1. 1-6). School principals accomplish those reductions in multiple ways and thus hold plural conceptions of a mean or median. Cognitive neuroscientists now corroborate Peirce’s propositions about perceptions of a manifold not only in the visual but also in the auditory and other realms of sensation (Seung & Lee, 2000), but remain undecided about whether neural receptors process such sensations in a two-dimensional or three-dimensional manner.

Saskatchewan school principals generally conceived of the average as a centre-of-balance. This conception is revealed not only in their commentary, but also may be inferred from their gestures and from the actions they described in response to particular averages. Their reasoning patterns with an average were abductively directed, for the most part, toward instruction for students who were below average within a local norm. The average as notation acquires significance insofar as it directs principals’ attention toward students who do not meet a shifting normative standard. The average functions as a prompt within school cycles for sustaining student, staff and school equilibrium as a continuously evolving state where contradictory interests and expectations must be continuously balanced.

Some school principals rejected the idea of a centre in student performance and questioned whether there is a typical student, raising a key issue in the relationship between organization and interpretation: do school principals autonomously hold a centre-of-balance conception, or do they only balance other people’s conceptions when managing the meaning of the average? If autonomously-held, then school principals bring their own intrinsically-evolving but pre-formed ideas of middle, grounded in previous experience, to a judgemental situation and thereafter manage from that position. This approach involves leading others toward a collective sense of middle within the school and community. If engaged in balancing others’ conceptions, the middle is an
ever-shifting point of conciliation or arbitration between others’ interpretations. When school principals balance other people’s conceptions of the average, then their sense of middle becomes fluid—an *ad hoc*, extrinsically-shaped, and ongoing social construction of individual staff members’ and perhaps parents’ views of centre—leading toward a mosaic of interpretations. At issue is whether school administrators rely on a personal, internal concept of the average when appraising student achievement, or whether their concepts of the average derive from the momentary positions of teachers and members of the public.

No clear-cut answer can be provided for this question. On one hand, nearly all principals indicated that they will intervene when the range of teacher-submitted marks contradicts their own sense of appropriate variation. Principals’ gestural depictions of a bell curve, their Cartesian diagrams of the statistic in relation to criterion measures, and their understanding of differences in the mean for the human and natural sciences would indicate a foundational understanding of these mathematical ideas and their application in judgements. School principals will ask pointed questions of the teacher, pair teachers to moderate their marks, and, infrequently, overturn a teacher’s assigned marks. Moreover, many principals will often talk to individual students when those students’ classroom marks do not match the principal’s expectations.

On the other hand, some principals describe the middle as a simple avoidance of extremes, preferring an ever-expanding and contracting zone through which they steer. Several findings from the quantitative phase of this study also suggest that school administrators do not have a resolute sense of the centre, and that they are particularly attuned to teachers’ concepts of the average and less influenced by the perspectives of individuals outside the school. These findings are: (1) expanding standard deviations after school administrators read the data display; (2) changing administrative beliefs to
become less rational-analytical; (3) changing administrative beliefs when information is confidential to the school rather than when it was made public.

Whether Peirce’s definition for a “conception” is terminologically equivalent to the cognitive psychologist’s “encoding in the mind” remains debatable. Certainly, Kahneman (1992) has distinguished between an anchor point (an intrinsically-held perspective in which are embedded one’s values), a point of reference (an extrinsic frame which does not represent the reader’s values), and an ambiguous normative standard (momentary, *ad hoc* activations of memory or constructs in the imagination) in negotiation and decision-making. Principals consistently described the average as a reference point or index. However, they were sensitive to its dynamic instability, recognizing that teachers assign differing weights to various assignments when they calculate students’ marks, and that students are constantly maturing and changing. The average as typical or representative student achievement may serve as a periodic and provisional anchor point when stakeholders consider student outcomes in a school plan. A fluctuating anchor has many implications for any social process of interpretation, because teachers and school principals will hold differing and unstable perspectives on the meaning and hence the value of any statistical notation, depending on how it is framed. A shifting reference point implies that there can be no consistent external bar against which to appraise student outcomes. An ambiguous normative standard implies that in formal bargaining processes, the parties may reach consensus, but hold very different meanings for the point of consensus.

When school principals explained their notions of average, they did not use mathematical or statistical terminology, but rather a more qualitative vernacular of colour19, clues, and warning flags. Principals’ metaphors often equated averages with

---

19 Peirce conducted original research in the sensation of colour, noting it: had trivalent qualities as hue, chroma, and luminosity; demonstrated mathematical bases; and could serve as the predicate of a
traffic signals, which give direction. The cues provided by statistics were a starting point for a conversation with staff, students or parents. Saskatchewan principals lead, for the most part, by asking questions of themselves and of others rather than by making pronouncements or drawing immediate conclusions. When seeking answers to questions about students’ averages, school principals will be guided by their own experience with the student, family, or staff members rather than by academic research. They concentrated on students’ classroom assessment marks rather than on externally-generated marks such as those from a provincial assessment, and they focused on gaps between student achievement and local standards rather than on gaps identified through equity principles such as gender, ethnicity, or socioeconomic status.

Both phases of this study indicated that when reading student achievement statistics, administrators focused not on average as typical student, but rather on those students whose achievement was below average. Several psychological explanations have been proposed for explaining this so-called negativity bias: survival instinct (Baumeister, Bratslavsky, Finkenauer & Vohs, 2001); loss aversion (Tversky & Kahneman, 1992); emotional or attitudinal responses to evaluation (Ito & Cacioppo, 2005); perceptual distortions caused by selective attention (Hilbig, 2012); and more complex information-processing (Rozin & Royzman, 2001). Cognitive psychologists have also posited a contradictory positivity bias: the Pollyanna principle upholds positive attributions as more significant in information processing (Matlin & Stang 1978; Mezulis, Abramson, Hyde, & Hankin, 2004; Peeters, 1971). However, classical pragmatism offers a more compelling, “effects-oriented” explanation for the negativity bias. The principal has legal responsibility for student learning and below average student achievement declarative statement. Even in an imperative proposition (for example, a stop sign), colour “would be general, since it leaves the interpreter free to give the other two coefficients of the color any determination that may suit him” (EP2. 394-397; Peirce, 1877).
scores might reflect badly on the principal’s reputation and the school. Of equal importance is the principal’s recognition that below average achievement scores foreclose educational and career opportunities for students.

Findings from both study phases also converge on principals’ tendency to categorize and thus simplify student achievement scores into three groups: those at average, above average, and below average. This seems a by-product of administrators’ reading processes—skimming and simplifying, rather than studying in depth. The triage allows principals to focus on potential problems, but leaves them prone to categorical overlaps. William James (1890) distinguished between the focus, the margin and the fringe in attention and in consciousness, suggesting that the margin controls meaning (pp. 254-260). When the principals accomplished a triage, the focus was consistently on below average student achievement, the marginal was average student achievement, and the fringe was above average achievement (Eriksen & Hoffman, 1972; Eriksen & St. James, 1986).

If the triage is intrinsic to school principals’ interpretive processes as a perceptual act of selection and classification when reading data, this presents a psychologically-based explanation for categorization processes at odds with that currently held in some policy circles. American researchers (Booher-Jennings, 2005; Youdell, 2004) have found that in high-stakes accountability systems, educators accomplish a triage by dividing students into safe cases, cases suitable for treatment, and hopeless cases, thereafter rationing resources to focus on those students most likely to improve a school's test scores. However in this Saskatchewan study, school principals conducted a remarkably different type of triage, focusing primarily on students with below average marks, calling in additional resources where possible to elevate their performance. Principals based their triage primarily on teachers’ classroom assessment scores, not on externally-mandated provincial or national tests which are not high-stakes and do not count in
Saskatchewan for student transcripts, promotions and graduation. The triage is intrinsic to school principals’ interpretation of the average and not simply a categorical scheme imposed through an external testing regime.

Furthermore, Saskatchewan school principals did not look at student achievement scores with the strategic goal of elevating “bubble students” (those students whose marks can be artificially inflated through rote drill to pass a minimal competency test). Rather, they sought to sustain a credible range of marks that reflected well on the teaching staff, on the principal and, ultimately, on the school. When Saskatchewan principals mentally and kinesthetically categorized students into three groups (below average, at average, and above average), the goal was not to ignore instruction for hopeless cases, but rather to elevate the performance of below average students, so these students perform at satisfactory levels.

In addition, both phases of this study demonstrated that school principals’ initial sense of the average derives primarily from their beliefs, rather than from external values that may be represented in the mathematical notation or in the statistician-author’s probabilistic ideas about statistical significance. Ryan (1999) has noted that many Saskatchewan educators approach student evaluation as a societally-assigned if somewhat uncomfortable, even distasteful professional obligation, which is value-laden. Educators view student evaluation as a moral obligation to the variety of constituencies they serve. Tensions are thereby created by the need to reconcile instructional processes with product (student) evaluations, by the mandate to meet curricular outcomes for mainstream students while recognizing the particular abilities of special education students, by the labour market’s need for capable employees, and by post-secondary institutions’ requirements for specific skills. In addition, student evaluation is influenced by the home and social situation of individual students, by the need to involve students in self-assessment, by the requirement to meet externally-defined standards, and by
contradictory concepts of fairness and equity found in what Ryan oversimplifies as “individual respect” versus “level playing field” viewpoints (p. 130).

When reading an average as a representation, therefore, many conflicting value positions may be invoked. School principals’ centre-of-balance conception may be an interpretive response to contradictory ideas about equity, which arise from assessments or testing, along with their statistical products. The location of an average in a normative distribution in a graphic can be implicitly defined as a problem in distributive justice. The underlying content in many discussions and controversies among partisan groups over averages often stem from irreconcilable beliefs about fairness and equity. Deutsch (1975) argues that fairness can be alternatively viewed as: treating all people according to their abilities, their efforts, or their accomplishments; attaining equitable outcomes proportional to inputs; fulfilling students’ needs; offering equal opportunity to compete without external favouritism or discrimination; meeting supply and demand in the marketplace; upholding the principle of reciprocity; meeting the requirements of the common good; ensuring that no one falls below a specified minimum; or considering all as equals in parity. These different value positions often conflict: “The most needy people may not be the most able, those who work the hardest may not accomplish the most, equal opportunity may not lead to equal reward, treating everyone as equals may not maximize the common good” (Deutsch, 1975, p. 140). Nor does any single value or subset of values have natural or universal priority.

Insofar as those Saskatchewan school principals I interviewed appear to have reflected on these issues, most interpretations of average student achievement indicated a values balance between equitable student outcomes proportional to school and student inputs, while also striving to ensure that no students fall below a curriculum-specified minimum, particularly at the elementary school level. However, school principals were especially aware that others will have contradictory values about an average as a student
grade. Therefore, a centre-of-balance conception may be the best of several vantage points from which to embark on the management of meaning. Upholding an elastic notion of centre, and asking questions rather than making assertions, enables a principal to manage values conflict both within the school and in the community.

In this way, a school administrator’s sense of professional and moral obligation is fused in a conception of the “neutral” average that may serve as a sort of internal compass which allows the principal to navigate around conflicting values to sustain equilibrium from within the organization. In that sense, the average becomes an external representation for the principal’s own internal moral ideas, allowing the principal to address conflicting values orientations both inside and outside the organization. Administrators adapt in their conceptions of a statistical average as centre, ignoring the different computational operations therein, to avoid or to elide the moral dilemmas inherent in evaluative practice. Two conditions appear to upset the principal’s sense of balance: a set of unfavourable student achievement results, as illustrated in this study’s second phase experiment; and staff members’ beliefs about the statistical average as socially-potent symbol, as illustrated in phase one interviews.

Among the alternate conceptions for the mean or median thus far identified in mathematics education research, the fair-share approach to appraising outcomes does not conform with principals’ circumstances, largely because several different averages are considered when allocating resources. Administrators consider PTR, teachers’ classroom marks, large-scale assessment averages, and class sizes. Principals’ commentary indicated that they did not view report card marks in terms of social justice principles but rather in terms of professional fairness or the adequacy of teachers’ classroom evaluation procedures. Professional considerations included having sufficient evidence to warrant a judgement, having clear criteria, knowing that teachers’ judgements are weighted and often subjective, and ensuring that attendance and other records together render the mark
as defensible. Nor is this a procedural, multiplicative (or add and divide) conception, given principals’ general disregard for formulae and the ready accessibility of software in schools. Nevertheless, school principals are aware of the cumulative and amplificatory effects of averages for individual students in their secondary and post-secondary trajectories.

It could be argued that the language of red flags and warning signs are examples of a signals-amidst-noise conception for the average. However, to claim as such would be to ignore school principals’ purposes. They do consider the average as a signal but not within a data set, nor in terms of a chaotic learning environment, nor as a manifestation of organizational disorder. Rather, a centre-of-balance conception better describes their periodic review of averages within systemic cycles. The average serves largely as a recurrent prompt to better prepare for impending conflict regarding expectations for or interpretations of student achievement.

In short, both perspective as a line of sight, and perception as our sensory impressions of a Sign, are related to school administrators’ beliefs as habits of mind—consistent with Peirce’s pragmatism and Prospect theoretic findings of a negativity bias in framed tasks. On one hand, Saskatchewan school principals’ perspectives for the statistical average are related to an organizational position which demands they hold multiple views of the mathematical concept as weighted and unweighted means, leading to a centre-of-balance conception. On the other hand, in school administrators’ perceptions of the statistical average, their judgements and hence their cognitive styles will subtly shift when faced with the prospect of negatively-framed information becoming available to teaching staff within the school. In both phases of this study, school principals’ attention was drawn to below average performance because their internal value systems demanded more flexible approaches, not because of external or public pressures.
8.3 Inferences and Inquiry

For Peirce, all thought consists of chains of inference, or reasoning with signs. As he contended:

The whole operation of reasoning begins with Abduction . . . . Its occasion is a surprise. That is, some belief, active or passive, formulated or unformulated, has been broken up. It may be in real experience, or it may equally be in pure mathematics, which has its marvels, as nature has. The mind seeks to bring the facts, as modified by the new discovery, into order; that is, to form a general conception embracing them. In some cases, it does this by an act of generalization. In other cases, no new law is suggested, but only a peculiar state of facts that “will explain” the surprising phenomenon; and a law already known is recognized as applicable to the suggested hypothesis, so that the phenomenon, under that assumption, would not be surprising but quite likely, or even would be a necessary result. This synthesis suggesting a new conception or hypothesis, is the Abduction. . . . This is not accepted as shown to be true, nor even probable in the technical sense,—i.e., not probably in such a sense that underwriters could safely make it the basis of business, however multitudinous the cases might be;—but it is shown to be likely, in the sense of being some sort of approach to the truth, in an indefinite sense. The conclusion is drawn in the interrogative mood . . . . This conclusion, which is the Interpretant of the Abduction, represents the Abduction to be a Symbol,—to convey a general concept of the truth,—but not to assert it in any measure (EP2. 287).

That overall reasoning process was repeatedly evident in interviews with Saskatchewan school principals as they interpreted the average. At its most general, abduction is a cognitive process for generating an assumption, an hypothesis, or a conjecture as an idea about something for which the person has declared an interest. However, abduction can take multiple forms (Gabbay & Woods, 2005), as Peirce recognized—and as evident in school principals’ read aloud interpretations of the statistical average in a variety of graphic displays. There were at least six distinct purposes for which a school principal would offer conjectures or preliminary diagnoses with the statistical average, all examples of the ampliative and defeasible reasoning that characterizes a pragmatic approach to educational issues. A conjecture may (1) provide an explanation for a given set of data or for some phenomenon. Or it can (2) facilitate the
generation of predictions, as evident in Phillip’s predictions about a student performance across several subject areas. Abduction may (3) permit the elimination of competing hypotheses, as Rick demonstrates through his analogy to the game of Clue. It may also (4) enable development of a more compact account, as evident in Louise’s triage of below, at, and above average students for further attention in her Watch List. Moreover, abduction may (5) permit the unification of disparate observations, as shown in Phillip’s consideration of attendance and performance in several subject areas. Abduction thereby (6) enables the school principal to steer around the average as an unimpeachable statement of truth or a morally-loaded idea, and to avoid a deductively conclusive statement that the statistical average corresponds with a classroom reality which must necessarily be changed.

Within the tenets of Prospect theory, however, such inferential processes are not directly investigated. For example, in Kahneman’s (2011) Thinking Fast, Slow, that synthesizes three decades’ of cognitive psychological research into thinking and decision-making, the words “reasoning,” “inferencing,” “induction,” and “deduction” are not mentioned once. Instead, the model of the mind which is upheld is mechanistic, equating thought to System 1 and System 2, with chapter titles such as “The Associative Machine” and the “Machine for Jumping to Conclusions.” Nevertheless, a pragmatic theory and a Prospect theory of interpretation share similar definitions for intuitive heuristics as simplified procedures for finding adequate but often imperfect answers to questions (Kahneman, 2011, p. 98). It is hard to avoid the conclusion that many cognitive illusions

---

20 Peirce was not averse to mechanistic views of initial perceptual judgements. One of his papers in 1900, “Our Senses as Reasoning Machines” (MS. 83), followed his 1887 article in the American Journal of Psychology which predicted computers with electronic wiring, digitization, a keyboard, and the use of symbolic logic (Peirce, 1887). However, he believed it was impossible to endow such machines with the genuine power of self-control and organic growth (EP2. 387).
which psychology has attributed to the illogical caprices of System 1 derive from the logical sequences in abductive reasoning that Peirce has described.

In school principals’ commentary, abduction appears as preliminary problem identification and a prelude to trouble-shooting. The statistical average provides a stimulus for questioning and formulating preliminary hypotheses, without an immediate need for definitive conclusions (Aliseda, 2005, 2006, 2009). These preliminary hypotheses are subsequently taken to the principals’ various constituencies for explanation, verification, or further elaboration. The logic is one of discovery for the principal, but may be viewed as the logic of justification by those who are queried. No clear distinction between the two can be drawn. These initial questions serve: as “bridge” between the conception of data and its effect within a school; between reading the data and acting on the result; between understanding the statistic and using it; and in the words of a more than one principal, between their office and the wider school community.

Therefore, it would be an exaggeration to say that school principals use the statistical average to make discrete or lapidary choices. Similarly, we should not over- loosely define the average in Dewey’s (1941) terms as a warrant for assertion or action. Saskatchewan principals often feel that a single average does not warrant a direct intervention requiring teachers to justify their professional judgements. Rather than frame the average as an immutable statement of truth, Saskatchewan school principals view a report card average and most statistical measures of central tendency as an approximation, as an incomplete and partial representation that may indicate issues lurking beyond. Nor should we adhere to an overly formal, syllogistic logic. Hunches and preliminary diagnoses are, at best, proto-propositions which the school principal formulates while reading statistics. Both statistic and hunch may be affirmed, elaborated or abandoned by the principal as her or his inquiry proceeds. The principals who
participated in this study reported using data as a foundation for questioning and indicated that they lead through interlocution rather than through direct intervention in the instructional process. We cannot say that data automatically spawns more data, as a form of decision-based evidence-making—except when teachers’ averages do not exhibit the range anticipated and the school principal asks for additional assignments to more accurately report a student’s standing. School principals do question teachers’ marks, and will call for more evidence if the marks appear inconsistent with the administrator’s sense of an acceptable distribution.

Therefore, we might be tempted to conclude that Saskatchewan school principals’ stance toward statistics is a form of transactional and not transformational leadership (Bass & Riggio, 2006). When discussing data, school principals evinced little interest in inspirational mission statements, nor expressed aspirations for achieving extraordinary outcomes. In this way, the interviews confirm Luo’s (2008) finding that principals consider data more useful for instructional and operational management than for pursuing transformational goals. Those upholding transformational leadership perspectives might dismiss principals as passively managing by exception, or as negatively highlighting incidents of poor teaching performance to the detriment of transforming school cultures and structures. Such a stance would ignore the tone or mood by which leaders ask questions. In contrast, transactional leadership is predicated on the notion that leaders must engage in daily exchanges with followers, in a continual give and take to effect change, but not with an attempt to barter or negotiate averages. What principals do is transform the statistic to a quality and a query with the aim of transforming learning for individual students, rather than stridently and immediately transform a school culture with statistics.
8.4 Administrative Beliefs and Doubt

In Peirce’s theory of interpretation, belief and doubt hold prominent, analogous but not identical positions to certainty and uncertainty within many contemporary cognitive psychological theories of decision-making. Whereas belief and doubt are cognitive states within a problem solver’s mind in a pragmatic view, certainty and uncertainty are exogenous conditions or variables under which decision-making occurs in Prospect theory (Tversky & Kahneman, 1992). The discordant terminology reflects differing assumptions about the reader. The pragmatic reader is endowed with purpose(s) which may be expressed in multiple ways. In apposition, the psychologically-compelled reader is often studied in controlled experiments without a detailed examination of individual motives, or with provision for multiple conduits of expression. As expressions of beliefs, habits have automatic or routinized attributes that straddle the cognitive, behavioural and emotive realms. Beliefs are states of comfort and satisfaction that may be disrupted by doubt; evidence contradicts or draws belief into question. Doubt is thus the spark of abduction in firing inquiry. Cognitive psychology has not yet considered doubt as a worthwhile zone for research. Organizational theorists have recently recognized Peircean doubt for its generative potential (Locke, Golden-Biddle & Feldman, 2008), even if doubt is not often considered a virtue either in leadership studies or in public communication. This study’s second phase MANCOVA analyses, with complex interactions of valence and audience, suggests that doubt may be an important link between the intuitive and experiential dimensions of System 1, and the rational and analytic operations of System 2. Rationality declined in statistically significant ways under all four framed conditions; the impact of framing on experientiality remains statistically ambiguous. Interviews with school principals likewise revealed that their doubts about the meaning of an average prompted them in school-based inquiry.
A key issue in evidence-informed decision making—in the classroom, courtroom or school—is whether doubt is reasonable and grounded in sufficient evidence, or more whimsical and grounded in personal fancy. Peirce argued that we reason from doubt occasioned by evidence toward belief and action, not from evidence toward doubt (EP1.114-115). Jurists and juries have long wrestled with a definition for reasonable doubt (Laudan, 2003, 2007; Mulrine, 1997), and the criterion standard of proof required for an explanation to be implausible or unbelievable. For Saskatchewan school principals, the criterion appeared as a point well below average when asked to read a set of student marks generated by teachers. This was particularly the case in the key subjects of literacy and numeracy, and especially if a below average trend was evident across multiple assessments. In the second phase experiment, principals demonstrated significant changes in belief not specifically according to a student’s below or at average standing, but rather when that standing was in combination with results labelled as confidential to the school. In other words, from the principals’ point of view, the standard of proof or the threshold of doubt appears to be set within an internal jury of educators rather than more broadly by an externally-framed court of public opinion. Although the second phase design did not predefine whether test scores were derived from a provincial large-scale assessment or were a provincial aggregation of teacher-constructed classroom grades, we can nevertheless say that substandard school performance provincially will create doubt, and prompt slight shifts in principals’ problem-solving beliefs to become less rational-analytic.

Saskatchewan school principals readily described the average as a measure of centre but, curiously, not of central tendency. They rejected the idea that an average represents a tendency within students and their achievement. Rather, as demonstrated in both phases of this study, principals themselves gravitated toward both a rational and an experiential centre in their beliefs when reading an average. In fact, their proclivity to
draw on intuition as heuristic, on experience, and on the heart may go some way to explaining why both elementary and secondary teachers will “pull” for students, giving them “the benefit of the doubt” in their marking practices (McMillan, 2001; McMillan, Myran, & Workman, 2002) as a compensatory variable in student evaluation algorithms.

American researchers have found classroom assessment practices, where:

low-ability students who tried hard would be given a passing grade even if the numerical grade were failure, although working below ability level did not affect the numerical grade. That is, an average or above-average student would get the grade earned, whereas a below-average student would obtain the benefit of the doubt if there were sufficient effort to justify it (McMillan, Myran & Workman, p. 205).

Such a trend may thus stem not only from teachers’ grading decisions, but also from school principals’ management of the meaning of marks with their professional staff.

Moreover, this study does obliquely indicate—through the different patterns in belief change stimulated by the different frames in this study’s second phase—that school administrators have ratio biases and difficulties with proportional reasoning. When 80 percent of students were identified as being at or below average, principals were more likely to shift in their problem-solving approaches than when 20 percent of students were identified as being above average. Although this study did not directly examine principals’ proportional reasoning patterns, three indirect observations may be made.

First, school principals are often asked to calculate (and held to account for) average pupil-teacher ratios and class sizes across the school and division. So we may presume they have a strong operational understanding of ratio and proportion, with an acute awareness of how these ratios translate into particular classroom configurations on September 30 each of each year. However, at different times during the year with shifting student enrollments and combinations of part-time and full-time teachers, principals will not have an exact knowledge of the actual ratios for a particular classroom: the number of teachers and students remain fluid with student flow. Second, principals repeatedly demonstrated in interviews that they do not ignore the denominator,
given its importance in budget allocations, class-size computation, and its direct
implications for teacher contracts and their own salaries. And third, principals uniformly
consider the averages provided by teachers through classroom assessment as weighted
averages, where various assignments and elements of an average mark receive
proportionately different weights by and among teachers depending on the subject and
year. We might conclude, therefore, that principals’ conceptions of proportion, ratio and
average become confused, not because of denominator neglect but rather because of a
nimble numerator that represents ever-shifting weights. Class overlap and inclusion
difficulties are not just a psychological phenomenon in categorization that explains ratio
problems; they are also a practical educational issue for school principals who must
discern who and what is computed in an average as a representation for both withinputs
and achievement outputs.

What is not customarily revealed within cognitive psychological approaches to
statistical meaning is that school principals’ conceptions of the average may be expressed
somaically, chromatically, pictorially, and not just arithmetically. Some might argue
that an unconscious gesture cannot be deemed a form of cognition or statistical
proposition at all. However, cognitive anthropologists have contested this notion, saying
the hands serve both depictive and conceptual functions. Both are “iconic” or “image-
related.” But whereas a gesture that represents an idea is made deliberately, if fleetingly,
by the gesturer in a conspicuous if not always conscious act of communication with
others, the gesture is also integral to conception of an idea. The manual expression is of
an idea seeking a form. Gestural manifestations of a conception are often unconscious, as
Peirce recognized—he pointed to the automaton-like responses of soldiers responding to
a drill sergeant’s order and the galvanic responses of a decapitated frog’s leg as
examples. Depiction and conceptualization then may be two different visual modes
which are distinguished by different frameworks of attention, the depictive somewhere
near the centre of visual attention, whereas the conceptual emerges from the periphery of attention, which proponents of embodied cognition argue are spontaneous “body thoughts” (Wilson, 2002).

This mixed-methods study thus raises profound questions about whether our concepts of the “average,” “centre,” “normality,” and “tendency” are treated separately and differentially in the mind in linguistic, mathematical, and/or geo-spatial terms. It also raises questions about whether these concepts are encoded separately or simultaneously in linguistic, pictorial or arithmetic forms, and whether they are imprinted psychologically or embedded physiologically. It would be valuable to know whether these concepts are melded together in the mind. Cognitive neuroscientists have recently argued that although motor and sensory processes are automatically engaged when people conceive and perceive, the brain still controls the body and semantics itself, not the converse (Mahon & Caramazza, 2008). Their grounded interaction framework, where the “activation of specific sensory and motor representations complements the generality and flexibility of ‘abstract’ and ‘symbolic’ conceptual representations” (p. 58) closely matches Peirce’s thinking, except in one respect. As with many cognitive science studies, the framework takes no account of the reader’s purpose(s) in conception.

If it remains difficult to parse out somatic demonstrations of the average from the cognitive in cause-effect relationships, it is likewise difficult to separate principals’ autonomous habits of thinking and acting from organizational cycles. These cycles include the annual budget development and staffing process, the review of marks before report cards are mailed, the development of school timetables and the assignment of teachers to particular groups of students. Recently, Spillane (2012) has distinguished between the ostensive aspects of institutional routines in data use (those formal organizational structures and committees in which data use occurs) and performance routines (the habitual behaviours of actors within institutional cycles). When viewing the
differences between a student’s average mark, a teacher’s classroom average, and a school-level average, principals may unconsciously adopt different conceptions of a centre in student achievement according to the demands of a particular school routine. The concept of the average thus changes depending on the situation and the purpose for which the mean, median, mode or other measure of centre is being considered. Principals considered the average as a weighted mean when considering a teacher’s marks, as an arithmetic mean when looking at large-scale assessment results, and as a distribution of marks that will be more important at end-of-term than at mid-term.

When asked, nearly all principals initially indicated they believed their school was at or above average. Many explanations have been proposed for this Lake Wobegone effect as it is known in American scholarship, or more evocatively as the Bullfrog effect in Great Britain—the tendency to inflate one’s achievements and capabilities, or to believe that all students and schools can be at or above average in their achievement. Such illusory superiority has been attributed to the self-serving impulses of ego (Kruger, 1999; Windschitl, Kruger & Sims, 2003), to an information theory mechanism which presumes a noisy conversion of objective evidence (observation) into subjective estimates as judgements (Hilbert, 2012), to faulty (gist) memory (Brown & Morley, 2007), to selective processes in choosing peers (Alicke, Klotz, Breitenbecher, Yurak, & Vredenburg, 1995; Dunning, Meyerowitz & Holzberg, 1989; Giladi & Klar, 2002), to a better than average heuristic which involves snap judgements based on ideal traits (Otten & Van der Pligt, 1966; Sedikides, Horton & Gregg, 2007; Sedikides & Strube, 1997), and to straightforward “big fish, small pond” processes of social comparison (Brown, 1986; Moore, 2007). Most explanations offered to date are psychologically-rather than logically-based.

However, a pragmatic explanation may also be offered for the Lake Wobegone effect. Peirce argued that deductive reasoning automatically invokes necessity, inductive
reasoning draws on ideas of probability, and abductive reasoning draws on expectancy (EP2. 233). Expectation is inherent to abductive logic: hope, and not a breach of regularity, animates our anticipations and explanations for events (EP2. 106-107; EP2. 88-89). Pragmatism accounts for illusory superiority, then, not by pointing to the psyche in a singular mind, but rather to the reasoning patterns and aspirational qualities within a collective mind. As Peirce noted, perceptual judgements at the abductive inception of thought are uncontrollable. Although moving small groups of students from the below average to the at average category might have minimal arithmetic effects on the overall group, we abductively expect that a few such successful actions will have a general effect for all students in a school and will elevate all to at and above average. School principals may boast of the above average qualities of their students as a motivational and morale-building strategy, as a public relations exercise to match community expectations, or as a way of elevating parental expectations in the home to improve achievement.

Disconfirming evidence is noticed, and is acted upon, but the belief that improving the performance of a few students will raise the average of the entire school defies straightforward arithmetic. In that sense, school principals’ beliefs are arithmetically uninformed if not irrational. The importance of elevated parental and teacher expectations for improved student academic performance is well documented (Rubie-Davies, Hattie & Hamilton, 2006; Seginer, 1983); expectation has been precisely defined and extensively studied as a measurable construct in both marketing and economics; the Pygmalion effect has been extensively studied in psychology. However, the management of expectations, as a facet of managing meaning, has not yet been seriously investigated by researchers in educational administration, perhaps because of the idealism brought to so many leadership archetypes.

The principals in this study, for the most part, did not see averages as tools for drafting or coercing teachers to act or behave in compliance with system goals. Scholars
have thus far emphasized two intentions that readers apparently bring to the manufacture and reading of statistics. Some researchers (Roth & McGinn, 1998; Smith, Best, Stubbs, Archibald, & Roberson-Nay, 2002; Smith, Best, Stubbs, Johnston & Archibald, 2000) emphasize inscription (to record such information so that it can normatively establish disciplinary authority) which has a bureaucratic function. Others (Roth, 2000, 2001, 2003) argue that graph reading is for conscription (to enlist and compel others into conformity with a dominant reading of events) as part of a scientific process in developing and achieving consensus about a positivist reality. However, when looking at Saskatchewan school principals’ reading of various averages, at least four other intentions can be identified: description (to depict and hence gather an awareness or understanding of student performance as a phenomenon within the school so that it becomes actionable); ascription (to ascertain those traits and attributes of students’ skills and competencies) so that better instructional arrangements may be made; prescription (to discover symptoms and diagnose shortcomings for eventual remediation of students’ perceived deficiencies); and reflection (to use the device as an imperfect mirror on one’s instructional or managerial practices). More importantly, school principals readily indicate that data displays which include averages serve as a platform for monitoring and hence managing instruction. As interrogative vehicles, data displays enable principals and potentially other professionals to ask more precise and focused questions about students, trends and existing organizational arrangements.

A more fundamental administrative question arises: are statistical averages considered primarily as precursors for problem identification and definition, or do they furnish the substantive bases for decision making in schools? We must recognize that Peirce’s triadic theory of interpretation was not formulated to explain decision making; nor was Kahneman’s and Tversky’s theory of dichotomous decision-making originally
developed to explain statistical interpretation. Whereas Peirce implies that averages as representations serve specific and limited but practical ends, Prospect theory implies that averages are a foundation for choice—that is, data-made rather than data-informed decisions. School principals, in their commentary, readily affirm that students’ average marks will shape if not determine those students’ destinies and, that grade level promotion in the late middle years and high school years in many ways hinges on report card grades. However, school principals more readily attributed choices made within and among courses to students’ and parents’ decision-making, rather than claim such decisions stem from their own administrative authority, school policy, or Ministry of Education regulation. In fact, secondary school principals often say choices are made piecemeal by students in a manner that reflects students’ academic diligence. Nevertheless, most principals are explicit in their commentary that they consider achievement averages when assigning teaching staff to particular classroom configurations of students and when allocating resources.

Perhaps the best way of clarifying whether data are used to problem solve or to decide may be by examining the triage, which posits two thresholds—that between below and at average, and that between at and above average. If decisions for groups of students are made at these thresholds, then the criteria at the thresholds should be clear and consistent. If decisions are not made but rather deferred, then the criteria will be slurred and inconsistent, with an emphasis on cumulative or cascading values rather than a single criterion threshold. Most principals did not articulate firm criteria at either threshold, preferring to retain an elastic or grey notion of the middle that shifts between subject areas, circumstantial demands, and temporal points in the school cycle. The conventional 50 percent criterion remains as the decision point in secondary schools, not
the class average as mean, median or mode. Thus, we would better characterize Saskatchewan principals’ use of the average as data-informed problem identification for instructional leadership, rather than data-driven decision making in student evaluation.

It is perhaps necessary to highlight what neither phase of this study revealed. No principal demonstrated much concern with the political symbolism of the average. Educational statistics, as Diane Stone (1997) emphasizes, often have symbolic value in policy-making. They function as metaphors by establishing likenesses between two things. Equating leadership prowess with above average basketball scores, or overall systemic failure with sub-Canadian performance on a single test are examples. Although Stone’s work is not predicated on Peirce’s semiotic nor on an understanding of statistics, she argues that counting becomes political because it requires decisions about categorizing, and about what or whom to count. The measurement if not the assessment of any phenomenon implicitly creates norms about too little, too much or just right. The average is thus one of those categorization devices that carries much symbolic cargo in political discourse.

In its ambiguity as a measure of quality or quantity, the average leaves substantial room for political struggles to control its interpretation. The mean or median may support the illusion that complex phenomena are simple, countable and precisely defined. Numbers such as a mean often provide a foundation for political communities because members share the same trait that has been measured, and because statistics can aid negotiation and compromise within that community. At the same time, by appearing to be precise, the statistical average augments the authority of those who calculate the statistic and those who embed the statistic in their rhetoric. Therefore, the numbers produced by an assessment become important instruments of strategic representation. Moreover, Stone (1997) emphasizes that synecdoches, as metaphoric figures of speech, are often used in political rhetoric to imply that a part of the policy problem represents
the whole. The average may therefore assume symbolic importance because policies based on examples are believed to represent a larger universe. Recent writers on educational micropolitics often draw on such metaphors in suggesting that the typical instances and prototypical cases are exemplary of the larger macropolitics of education, or vice versa (Blase, 2005). In first phase interviews, school principals sometimes expressed skepticism about the macropolitical uses of numbers; however, they readily reported intervening with teachers, students and parents in the micropolitical realm, not only to manage learning for at risk students, but also to manage the meaning of report card grades.

School principals’ agency with the average is expressed in two ways: in the type and range of questions they ask; and in the retroductive pursuit of explanation and adjustment of previous learning conditions for students by holding the average provisionally as consequent, not cause. The principals’ overarching purpose is expressed in a centre-of-balance conception, implying that equilibrium is the desired end. Econometricians have proposed three properties of an equilibrium state (Dixon, 1990): when the behaviour of agents is consistent, when no agent has an incentive to change its behaviour, and when the end state is the outcome of some dynamic process. All three properties hold at the present time for those Saskatchewan school principals who were interviewed. A system is considered stable, asymptotically-speaking, if it always returns to the same point after small disturbances. If the system moves away from the equilibrium after small disturbances, then the equilibrium is unstable. Stability is a key notion in the formal sciences because it refers to what we call "reality:" the object must have some degree of regularity to be observable.

Yet we would be mistaken to situate student outcomes registered as normative statistics as public goods within neoclassical economic models of either partial or general equilibrium theory (Greenwald, Hedges & Laine, 1996; Hanushek, 1997; Levačić &
Vignoles, 2002; Rolle, 2004; Todd & Wolpin, 2003; Verstegen & King, 1998). School principals do not negotiate and elevate test scores gradually through some managerial process akin to a Walrasian tâtonnement in supply and demand markets. In a school, there are just too many countervailing demands on a school principal’s time and too many fluctuations in educational resources to draw a linear and rational line between inputs and outputs. If anything, this study demonstrates that school principals’ and teachers’ withinputs substantially alter conceptions and expressions of outputs in the form of classroom grades.

Indeed, school principals’ interpretations and behaviour with the average might be better modelled within managerial economics as a problem in satisficing, not optimizing. In economics, satisficing is a behaviour that involves a search strategy, conflicting goals, an emphasis on putting out fires, and adaptive aspirational levels. Satisficing is an attempt to achieve at least some minimum level of a particular variable, but does not necessarily maximize an econometrically-construed value (Radner, 1975, 2005). Of course, such behaviour falls within the bounds of Herbert Simon’s (Forester, 1984) bounded rationality, reflecting a concern not with Pareto optimality but rather with meeting a standard of remediation.

Likewise, we might remain skeptical that average student achievement serves as an unbiased or singular conversion index within an overarching political systems theory (Easton, 1957, 1979) in the ongoing input-output exchange between the school as political organization with its environment. School principals do explicitly recognize statistical inputs as demands from the community (as expressed, for example by survey opinion) and emphasize that supports in the form of operating grants from government provide the raw material or information that the school must distribute. The allocation of grants and other resources such as staff expertise and supports do permit the school principal to perform her or his work of satisfying community demands—principals are
explicit in this regard. And the principals interviewed do take seriously their role in concert with teachers to authoritatively and normatively allocate highly valued, societal “objects”—authoritative appraisals and decisions about a student’s potential for citizenship, the labour force, or for post-secondary education, as represented by grades. Many principals do recognize that the characteristic outputs of any school, as a political system, are decisions about a student’s level of competence, skill or ability. In addition, most principals described the demands for funds, staff time, and staff attention that originate from within the school itself. That is, teachers and principals are performing political roles within the organization when allocating resources.

If the school is defined as an organization vested with the statutory responsibility for making authoritative decisions as outputs (Easton, 1957), then Saskatchewan school principals do not assert their normative authority directly over student marks as statistical representations of such decisions. Rather, they manage the meaning of such statistical notations more obliquely within organizational cycles by querying teachers and students as organizational members about the sufficiency and credibility of the evidence that has been assembled that would justify professional choices. These data-informed decisions are not binary as upheld in Prospect theory, but invariably occur on a scale or continuum, as implied in classical pragmatism. Internal rather than external demands are considered to be more pressing: from the school principals’ perspective, a supportive state of mind among teachers as “a deep seated set of attitudes or pre-dispositions, or a readiness to act on behalf of some other person” (p. 390) appears to predominate over external demands made by the public. Indeed, principals’ commentary denies rather than affirms political interpretations for an average. They see their role as modulating internal organizational demands before, not as these demands become political issues. Saskatchewan school principals therefore aim to forestall rather than abet politicization of student averages.
If anything, school principals’ own interpretations of the statistical average are rejections, not applications of political ideas.

This satisficing approach may help explain the conservatism of school principals in decision making under conditions of risk, as compared with their business counterparts (Brown, 1970). School principals are more risk averse than private sector managers, hugging the centre, acting directly and individually to elevate the achievement of below average students, and not trying to maximize the performance of all organizational members. Brown speculated that public sector managers must conform to social norms more so than private sector managers who value quantitative feedback and performance measures as indicators of entrepreneurial success. Saskatchewan school principals in this study demonstrated through their focus on the below average student that they value the attainment of minimal outcomes more than do private sector managers who value improvement of outcomes. The public civil servant is more conservative than the private sector manager because of conflicting socio-educational purposes (Lewis, 1990). This becomes a contentious issue when policy makers impose Total Quality Management approaches to excellence with singular aims—an exercise in optimization—on school principals who habitually juggle multiple and often contradictory public goals—an exercise in satisficing.

8.5 Leadership and Leading Questions

School principals tend to ask questions rather than make assertions when invited to interpret a data display. There are several possible explanations for this behaviour. A century of educational research in North American schools indicates that teachers will ask between 300-400 questions in a typical working day; constructing and responding to verbal and written questions comprises up to 80 percent of a typical teacher’s professional workload (Wilen, 1982). Although there was an intensive effort three
decades ago to better understand teachers’ questioning techniques (Carlsen, 1991; Gall, 1970; Harrop & Swinson, 2003; Samson, Strykowski, Weinstein & Walberg, 1987), no researchers have pursued the purposes of questioning by school principals. In the absence of research in this realm, we may only speculate. Peirce argued that any abductive question reflects “genuine doubt” about meaning. As Rick noted, the data display does not yet establish a complete or compelling fact. As a representation, the average remains only a sign, and the school administrator still has to assemble a wider array of information for its actual or actionable meaning to be determined.

Many other reasons why school principals’ first impulse is to ask questions may be inferred from their commentary:

- A *role-carryover* hypothesis implies that questions come habitually, perhaps naturally, to school principals who have taught, and in some cases continue to teach part-time, concurrent with carrying out their administrative responsibilities. As one principal noted, they “wear two hats” and thus can adopt two alternate perspectives on statistical meaning, which are mediated through questions.

- A *power modulation* hypothesis revolves around the relationship between authority and the average grade. Principals ask questions to moderate and “dial down” the intensity of a potentially contentious topic. The principals interviewed for this study repeatedly referred to the average as “a conversation starter.” Questioning is a way to sustain a collegial if not exactly a co-equal relationship within a First-Among-Professional-Equals orientation to instructional leadership, or a way of initiating a dialogue with parents. Asking a question is thus less an artificially-contrived collegiality for addressing professionally-incendiary material, and more a form of professional etiquette for dealing with divisive topics. The question functions as a velvet glove for assertion (Koshik, 2005).

- As a *transfer-of-responsibility-tactic*, school principals’ questioning is a way of encouraging teachers or others to accept ownership for results, rather than assuming that the principal alone is responsible.

- An *accountability-anticipation* hypothesis suggests that principals forecast that negative test results will elicit phone calls from parents, or negative messages from others. The principal asks questions to gather more information so that s/he can be prepared for complaints from parents or others and provide parents with more complete information about students’ learning or instructional circumstances.
• The *managerial margin-of-maneuver* explanation assumes that questioning upholds school principals’ autonomy. Unlike statements or assertions which intrinsically require the school principal to stake out ground and take a position, questions put the onus on the respondent for answers, thereby affording the principal some degree of freedom to change positions or shift in his or her interpretation of the average. The gaps between various means are thus the location wherein the school principal finds room to negotiate.

• A *decision-deferral* hypothesis suggests questioning is a way of delaying and defusing potentially contentious issues, allowing them to fade or enabling rival interpretations to moderate through either the lapse of time or subsequent dialogue.

• An *innumeracy-compensation* explanation suggests that because principals may have neither a sophisticated understanding of mathematics nor comfort with statistical matters, they call on people with more mathematical knowledge for assistance in interpreting a data display. A question based on a hunch or conjecture may be a genuine request for help in overcoming shortcomings in their own statistical knowledge.

• A related, *bridge-off-the-island hypothesis* implies that the principalship is often a lonely and isolating position. Questions enable the principal not only to ascertain the facts, and gain a more sophisticated understanding of the meaning of an average, but also to sustain an ongoing conversation with teachers, seeking their assistance with important educational issues. Questions are vehicles for sustaining a network of working relationships with staff members and parents.

• The *student grade-as-weighted average* explanation suggests that school principals will not be able to grasp the meaning of average student achievement without querying teachers who often have idiosyncratic algorithms for computing a grade.

• The *statistic-as-narrative-constituent* hypothesis suggests that adults are at heart, story-tellers. Accountability narratives are actually oral stories with embedded numbers, as semantic tableaux. A question is a way of starting the construction of such stories with statistics.

• A *critical stance* hypothesis suggests that such questioning enables the principal to distance her- or himself from and retain a critical posture toward test statistics, while still fulfilling administrative obligations. The principal does not have to either personally or positionally invest in the statistic, but yet still can report to superiors that s/he has acted upon the information by interrogating subordinates, while still retaining the latitude to express skepticism about the validity of the finding and disavow the result.

• The *professional growth* hypothesis implies that statistics serve as mirroring devices, which enable educators to reflect on their teaching, instruction and students. In this line of thinking, the contemplation of averages becomes an ongoing form of inservice. Questioning is a form of coaching, for which some principals may have received training.
An impression management explanation proposes that leaders’ questions are important vehicles for reinforcing or upholding desirable social images or identities as professional educators. Leading questions are not necessarily duplicitous in their intent, nor social-psychologically distinct from cognitive psychological phenomena (Tetlock & Manstead, 1985), but are useful for portraying oneself before others in culturally-desirable ways (i.e., non-adversarial, insightful, helpful).

Confirming the foregoing hypotheses may require drawing on concepts and assumptions from several academic disciplines (Yang, 2006), but social psychological or symbolic interactionist research designs appear the most promising. Social psychological approaches also appear congenial because of the strong spatial connotations in principals’ commentary about the meaning of an average. Psychologists and anthropologists have long recognized that people have a sense of personal space through proxemics and kinesthetics (Hall, 1974) that conveys meaning. Curriculum theorists also postulate a pedagogical space, distinct from but often confused with a sense of place, for discussing an issue (Gruenewald, 2003). We might also note that a psychological sense of social space has both horizontal and vertical dimensions that interact with both verbal and nonverbal behaviour (Hall, Coats, & LeBeau, 2005).

Finding a spatial or locational connotation in principals’ mindset may in itself be unsurprising, given school administrators’ customary responsibilities for timetabling and facilities management. Managing the meaning of an average overlaps with the management of plant, people and program as traditional administrative tasks. The term “direction” itself implies space. Whether the question and ensuing conversation occurs privately in the principal’s office, with parents in an adjacent boardroom or conference room, in the hallway in passing, one-on-one in the teacher’s classroom in a formal meeting, or as a formal agenda item in a staff meeting, are deemed crucial to how others will interpret its significance. Each location is invested with different kinds of authority, calling forth alternative senses of territorial imperative. Even if we know little about how
school leaders use space and place formally and informally with statistics, the astute
manager of meaning clearly has at her or his disposal a variety of registers to deliberately
modify the significance of a report card grade or the results from a large-scale
assessment.

8.6 The Average as Quality and as Quantity

The meaning of a mean is malleable because a statistical Sign may nearly
simultaneously express either quality or quantity. In interviews, many principals said that
the statistical average as quantity impedes appropriate communication of student
performance as quality to parents and students, and that various statistical measures of
performance interfere with students’ development because they are a judgement that
affects students’ self-esteem. They also said that a percentage average on a report card
becomes pedagogically meaningless without accompanying task or criterion information.
Interviews with principals revealed that they categorized averages according to
provenance (who produces the average) rather than according to the averages’ actual
mathematical properties (the types of averages that may be produced). This suggests that
case studies rather than variable-oriented research will be more meaningful in leadership
preparation (Greenhalgh, 2007; Ragin, 2004) for both principals and vice-principals, who
share similar dispositions when reading statistics.

In addition, principals tended to describe averages in terms of colour or traffic
metaphors as a gloss applied directly to the numeric icon. Obviously, as Craig noted, it
makes a difference whether yellow signifies average student achievement or achievement
that is marginally below average, or whether a black and white numeric figure is
reformatted as red and green space in a graphic. This points not to the need for advanced
training in software formatting options for more colourful graphics, nor the design of
elaborate information systems (Stamper, Liu, Hafkamp & Ades, 2000), but rather to more extensive orientation to the ethics of data presentation.

Likewise, the randomized experiment demonstrated that administrators’ conceptions of average student achievement as quality interact with their perceptions of the average as quantity. Statistically significant interactions between audience and valence lead to small but discernible shifts in problem-solving dispositions followed school principals’ reading of reports with alternate but equivalent ratios. It is worth emphasizing that the four frames were identical in prose as criterion statements of student quality, in their graphics as pictorial displays, in their accompanying icons for percentage and probability, and in their assignment of black and blue colours. Moreover, the sources for the average, and the planning situations in which the average was to be considered, were specified in identical prose. All four frames were temporally located in the conditional future tense, not the past, to minimize the impact of memory. In short, all qualitative dimensions were held constant, including the ordinal properties. The only variables adjusted were quantitative descriptions of proportion: an “80 percent” or a “20 percent” as cardinal properties, and “below/at” or ”above” as nominal features. The upshot was a statistically significant decline in school administrators’ rationality, contrary to a key proposition in Prospect theory. However, school administrators in all four groups sustained a qualitative proclivity for listening to the heart, for being guided by gut instinct, and for relying on hunches.

In fact, both phases of this study would draw some precepts in Prospect theory and Fuzzy Trace theory into question. Although this study’s phase two experiment supports Prospect theory in its psychophysical propositions about perceived gain and loss, the “rational-analytic self” did not over-ride the “intuitive-experiential self,” nor vice versa, in a statistical reading and planning situation. Prospect theoreticians have
lamented the absence of a formal framing theory (Tversky & Kahneman, 1992), although some have attempted to develop one within a psychology of expectations (Nelson, Oxley & Clawson, 1997). Peirce’s semiotic may fill that void: researchers are simply making minute adjustments to a Sign or its environment for an identical object, to see how interpretation changes.

At the same time, a key Fuzzy Trace theoretic proposition about the operation of memory may be inverted. Rather than seeing the average as a qualitative gist or husk stored in long-term memory as a simple more/less/equal which interferes with sophisticated or verbatim statistical representations in working memory, the converse may be true. The qualitative average itself may be a verbally, pictorially and chromatically complex idea in working memory which, when projected into the future, interacts with a mathematically reductive, logically summative, and numerically austere representation of the average stored in long-term memory. In other words, the numeric kernel of an average as a quantitative point within a data set, the simplified add-and-divide algorithm memorized in school, the precisely-etched image of a single Gaussian curve which is upheld in undergraduate statistics courses, or the thin skeleton of numbers on a school transcript could be the non-fuzzy substance of long-term memory, as a quantitative and not qualitative gist. Although we cannot yet discount Fuzzy Trace theorists’ ideas without explicitly-timed and carefully-designed experiments in working memory, repeated-measures studies could yet affirm and integrate propositions within a unified Fuzzy/ Prospect theory, for which Brainerd and Reyna have recently argued (2011). Indeed, Peirce’s 1892 article called “The Law of Mind” forecasts a reverse and not similar directionality in perceptual intensity (p. 550) when interpreting events retrospectively and prospectively, consistent with the synechism in his pragmatic philosophy.
8.7 Practical and Statistical Significance

Peirce devoted much of his writing to issues relating to the meaning of signs. Quite clearly, there is a difference between conceiving and reasoning within the bounds of a data set, and conceiving and reasoning from a data point in a relatively unbounded environment. Scholars have long wrestled with the average and its meaning. Aristotle identified two kinds of means, one revolving around the location of the object, and the other around the location of the observer. Both types of means concern the continuous and divisible. When the average is considered as an object, the observer can take more, less, or an equal amount according to an exact midpoint: the concepts of more, less, and equal are set in terms of the object itself. Under this objective definition, the arithmetical mean is equidistant from the extremes and therefore identical for all. The average of six between two and ten is an example. In a practical sense of the average, however, these concepts are relative to us as observers. The middle is a general intermediate zone between excess and deficit. Under the “relative to us” as observers’ definition, the mean may not be equidistant from the extremes but rather “neither too much nor too little, nor the same for all.” Here, then, the extremes admit of different measures of centre, where observers seek a midpoint among themselves rather than within what is observed.

For example, from four observers’ vantage points standing at each corner of a square room (presuming from the beginning that the four observers agree they are looking at a table top as object), the centre point of the table may appear mathematically and geometrically identical. But for people in a rectangular room where some observers are further away from or at lower levels than others, those seated distant or lower than the table top have a different perspective on the midpoint than would people who are close or above the table top. In determinations of average student achievement, school principals not only recognize that people have different vantage points, but often no consensual
understanding of what achievement actually is. This definition of the mean “relative to
us” has both epistemological and ontological implications. Knowledge of the
intermediate must consider differences between observers; ontologically, the intermediate
itself may not be identical for everyone, but depends on an inter-subjectively agreed
definition for both object and interpretation of the middle.

Each perspective demands a different mode of reasoning. Determining statistical
significance is primarily inductive and deductive in its patterns, whereas practical
significance is abductive, requiring a leap in thinking to understand others’ perspectives.
The deductive is inherently convergent and analytic, whereas the abductive is synthetic
and divergent in its attribution of antecedent and consequent relations. Data analysis
therefore makes different demands than statistical interpretation. Data analysis revolves
around means or between location and outlier, whereas statistical interpretation involves
reasoning among object, sign and interpretive effect. In data analysis, one’s thinking is
bound by the data set of observations, but in statistical application, boundaries are
assigned in a more open manner through reader purpose. It follows that the directionality
of our thought becomes important, through looking at causal sequences. In data analyses,
the causal sequences are affixed through the formal design of inquiry, whereas in
practical interpretation, such sequences are circumscribed only by individual or collective
belief. Nevertheless, it remains impossible to draw sharp lines between the three modes
of inference. As Peirce recognized, there is a continuum between inductive and
abductive reasoning methods. This continuum derives from perception, and from the
multiplicity and inherent vagueness of many objects and their attributes which are
represented in any Sign. That is, any average as construction must embrace a variety of
attributes of student achievement, but can only partially do so in measurement.
Statistical significance and practical significance thus rely on different types of inference. In both, the construct of interest is assessed, if not measured. Samples are used in both to discern some feature of a population or process. Both approaches aim to detect differences, for example between hypothetical (expected, targeted) and the actual measures of the population or process. In school principals’ ordinary encounters with a statistical average, however, the central purpose is to detect trends in processes such as shifting means in student work—whether the grades assigned by teachers exceed expected ranges in policy or leadership preference, or whether there is an impending disruption in a school’s equilibrium. The equivalent of a null hypothesis in formal science becomes “the educational process is stable” in educational administration, where there is no “significant” difference between teachers’ expectations and the individual student’s performance. Within the formal language of experimentation, this implies there is only common cause variation: the mean and variability do not change much. The alternative hypothesis would then be “there is a change in the process” with a special cause, but school administrators must leave unspecified what this instructional, home or personal cause might be when beginning to read a graphic display. The pursuit of that special cause becomes the grist or gist of abductive reasoning after reading the data display; it is part of and does not precede the interpretation.

When interpreting a distribution of student scores, school principals generally ignored the footnoted notations at the bottom of the page that indicated probability: none could describe the meaning of a $p$ value and the $< .001$ at the bottom of the page. It might be argued that this is strong evidence that principals have limited statistical understanding and an inability to appropriately interpret most forms of quantitative research (Makar, Bakker & Ben-Zvi, 2011). Conversely, it might reflect that principals have informal conceptions of probability in practice, as a Bayesian degree of belief or a
subjectivist degree of certainty. If this latter situation is the case, principals’ degree of conviction associated with a number would be logarithmically proportional to its distance from the mean set at 0, using Dempster-Schaffer theoretic rules (Huber & Schmidt-Petri, 2009; Schafer, 1990) to include judgements about the reliability of the evidential source. That approximates what Peirce called the conceptualist view of probability (EP1. 142-161; Peirce, 1978).

What appears to happen is that principals’ intentions are rendered more precise when they are considering average student achievement. Specific equity concerns and particular educational deficiencies are brought into sharper relief for individual students as “human figures,” using the statistical representation of the average student as ground. This is the opposite of putting students in the background, and using them to clarify our statistical ideas. Stated another way, the statistical average is the backdrop against which we more clearly see the precise features of the exceptional. This corroborates the Gestalt perspective which Peirce upheld (EP2. 124) wherein the background whole calls out the parts, rather than the figurative parts constituting the whole. In fact, anthropologists who have studied adolescent’s conceptions of mean and median argue that such ideas support emerging goals (Nasir, 2000; Saxe, 1991) in individual cognition.

Hence, in their dispositions and actions with the average, Saskatchewan school principals demonstrate that the practical significance of a key statistic is epistemologically different than the statistical significance. Whereas practical significance may point toward multiple paths in inquiry, statistical significance is most often measured along a single line when comparing two means within a data set. If statistical significance is an affair of efficient causation, practical significance relates to final causation, engaging the ideals and purposes of the reader. Whereas statistical significance depends on the formal application of mathematical procedures to a collection
of observations in a data set, practical significance must be based on a coherent theory of signs. Practical significance depends on the antecedent-consequent reasoning of the participant, whereas statistical significance draws on the observer’s views of cause and effect. Peirce’s brand of pragmatism offers a comprehensive theory of interpretation that straddles the two paradigms, in contrast to contemporary cognitive psychology. The practical significance of a statistic for leading changes in belief is not equivalent to the statistical significance of a change in leaders’ beliefs.

What then are appropriate criteria for evaluating practical significance? For statisticians, the notion of an effect size has gained sway over the past decade in social science research, in the wake of intense debates about the validity of null hypothesis significance testing (Cohen, 1994; Johnson, 1999; Schmidt, 1996; Thompson, 1996). Here again, there are a variety of formulae in psychometrics ranging from Cohen’s $d$ to partial eta squared. However, from the point of view of the reader, the significance of any statistic will be found outside the bounds of a data set. Regardless of which measure of practical effects is chosen, effect sizes are indices computed through the application of mathematical rules to a collection of observations. Peirce insisted that we acquire new ideas from direct experience, not from introspection alone, nor from application of the deductive axioms of mathematics. The statistician-author’s attempt to quantify practical significance as an effect size is an appeal to other statistical authors. Furthermore, Peirce insisted that the triangular relationship between sign, object and interpretation is not severable. Thus, qualitative methods may be preferred if not prerequisite to ascertaining practical significance, and we are left with plausibility as the best if not sole criterion for appraising the significance of principals’ conceptions for average student achievement.

How significant, then, are statistical representations of average student achievement for Saskatchewan school principals in their instructional leadership? Those
arguing that averages do not appear particularly significant in school administrators` practice will point out that the differences in means and effect sizes observed in the experimental phase of this study are small. Replication is required. Moreover, they will point to qualitative testimony which indicates that reading a statistical report of student achievement has minimal impact in shaping principals’ outlooks. It simply spurs principals to ask questions of parents, students and teachers without taking decisive action. Few transformative changes in school structures are apparent; nor are administrative cycles realigned to improve student achievement. Rather than leading a transformation of schooling with bold mission statements and major program changes to achieve the extraordinary, school principals manage by exception with questions to fulfill ordinary expectations in predominately laissez-faire approaches with teachers. Principals report that they will infrequently intervene with the at average and above average student, whose performance must be elevated to change a school’s performance, attending instead to proportionally small numbers of students who do not meet the fluctuating and transitory standards of means which are computed in an ad hoc and inconsistent manner by teachers. Saskatchewan school administrators’ overall managerial intentions appear to revolve around correcting temporary imbalances and sustaining a pre-existing stability in student achievement within a narrow band of subjects.

Moreover, skeptics about the significance of average student achievement for school principals will note that reading a student achievement report has relatively small effect sizes on school administrators’ problem-solving dispositions. School administrators’ pre-dispositions to equally draw on intuitive and rational approaches when reading an achievement report remain durable overall, regardless of the public or negative frames supplied by an outsider. The average student remains marginal in school principals’ attention, whereas the below average student evokes largely abductive
reasoning which some scholars argue is ignorance-preserving (Hoffman, 1999), rather than knowledge-generating. Others argue that abduction is not logical at all (Plutynski, 2011), given the absence of a rule in thought. Troubleshooting, detective work, and problem solving are aspects of any managerial orientation that strives to conserve the organizational status quo. In short, although principals’ reading of averages within a normative distribution may register probabilistically significant effects on school administrators’ habits, the practical significance of an achievement report is minimal in altering principals’ prior dispositions. Those who would argue that abduction is not logical would conclude that pre-existing administrative beliefs trump the evidence found in test outcomes.

Not so, proponents of statistical use will counter. Those arguing that the average has significance will point to the many practical implications of reading an average that are apparent in school principals’ commentary and behaviour. Notions of the average have been assimilated to such a degree that principals will unconsciously demonstrate their conceptions in gesture. Gestures are widely recognized in communication studies as especially influential in leadership and persuasion, helping to define organizational norms and directions (Streeck, 2009). Any educator will point to the importance of questioning in cognitive development; most researchers will emphasize the pertinence and precision of a question for shaping inquiry and hence knowledge generation. Those influences on principals’ thinking are often deemed the central concern in both leadership preparation programs and the apprenticeship of vice-principals within schools. Even if the effect sizes were small in the second phase of this study, we must not forget that even small changes in resilient elements of the personality—particularly that of school principals, whose reputation is predicated on consistent, solid, common sense—will have substantive
implications. This is especially true if the “treatment” is relatively mild in its dosage adjustments and momentary in its application. We might anticipate more pronounced changes when actual student achievement results are presented to a school, and particularly if repeated over a long run. The doubts that negatively-framed information evoked in this study appeared sufficient to occasion a statistically-significant deterioration in school administrators’ rationality as defined within some realms of cognitive psychology. Even in the artificial circumstances of a controlled experiment, small manipulations in the frames for an average within a normative distribution of scores seemed to have detectible effects in changing dispositions. Moreover, statistical self-efficacy was a strong predictor of school administrators’ adoption of rational-analytical cognitive styles. Perhaps more importantly, we must remember that an equilibrium state is not synonymous with stasis. Averages may not dramatically perturb either the school or its leader, but they have discernible effects in dynamic systems such as a school. The location of an average within a normative distribution may not trump belief, a proponent of practical significance would argue, but it may discernibly alter belief.

To adjudicate between these contradictory positions on the significance of statistics—one upholding the statistician-author’s expertise and the other upholding the reader’s experience—we might carefully scrutinize Peirce’s writings. In a letter to William James, Peirce defined the word “significance” (EP2.494-498) thoroughly and precisely within his pragmatic philosophy. Significance resides in the questioner’s “purpose for asking a question, and what effect its answer would have in her plans for the ensuing day” (EP2. 498): “Her Sense is the Impression made or normally to be made. Her Meaning is what is intended, its purpose. Her Significance is the real upshot” (EP2.
These three levels correspond to “maybe’s,” “actualities,” and “would be’s” (EP2. 500-501). That is, the significance of a statistical average resides not in the social context, nor in a comparison of two means, nor in the notation as Sign, nor in probabilistic ideas of chance. Rather, the significance of an average resides within the reader-as-inquirer’s purposes in light of prospective changes in her or his own behaviour. For the research community, the implication is we should look for significance not in the statistician-author’s findings within a journal article, with its notations of $p$ values and calculations of effect size, but rather in the publication’s introduction, where the researcher announces his or her aims, research questions, and the anticipated benefits of the publication.

For the school principal as a reader of statistics, the significance of an average for instructional leadership involves auto-reflecting on his or her own purposes before asking questions of others about the statistical average, and considering the effects in terms of anticipated changes within her or his own managerial practices. If interpretation is a free walk on firm ground, the ground is leadership purpose, not the statistic itself. Averages do not offer unequivocal solutions, but only enable more precise questions to be asked. Thus, pragmatism upholds the rights of the statistical-reader rather than of the statistical-author to determine significance in a final interpretant (EP2. 498-499). Whether defined as “the proper significant effect,” “the intellectual purport,” or “the signification” of a sign—as alternative terms in Peirce’s lexicon—all revolve around the reader’s purposes, not those of a statistician-author and his or her notations.

---

21 William James believed that perceptions or first impressions had meaning, but Peirce did not (EP2. 401; EP2. 419; EP2. 449-450).
We can therefore conclude that average students and their achievements as typical or representative, per se, remain at the periphery of Saskatchewan school administrators’ attention. School principals do periodically monitor student marks, and do consider the average from several different sources as one among several indicators that the school is operating smoothly. However, they do not habitually consider the qualities of the average student as an iconic representation, except insofar as the average student is background for highlighting those below or well above a fluctuating notion of typicality. Average budget allocations and the equitable distribution of resources are considered to be more “real” than statistical constructions of their students. In their disregard for probability values as frequentist statements of chance, no less than in the marginal attention to the average as typical student, the mean has little statistical significance as symbol for the principal. However, in their focused attention to those students not achieving a flexible normative standard as represented in a numeric average as index, school principals do act, and do pose searching questions—which educators recognize as among the most powerful instruments in pedagogy and inquiry. Below average has substantive practical and statistical significance, regardless of the measure of central tendency under consideration.

All the foregoing argues compellingly that school principals will not remain locked into a mono-logical and uniform mindset, let alone stay committed to pre-defined actions in relation to representations of an average. The average as synecdoche is complementary to, not the antithesis of pluralism. Over 50 years ago, one of the first North American researchers (Briggs, 1964) into high school classroom grades highlighted dialectical extremes between perspectives on the average as tool, or as tyrant. That polarization obscures much. As William James (1909) recognized in his essays and Charles Peirce noted in his description of infinite Dedekind cuts in a continuous number
line (RLT. 145-164), the world is inescapably multiple (CP4. 445-446), even for this most normative of statistical representations and for this most deceptively simple of arithmetic signs. Saskatchewan school administrators’ diverse interpretations are a salutary reminder that our own ideas of the arithmetic average, which seem so obvious and uncomplicated, are unlikely to appear the same way to others next to us in the same room.

8.8 Reflections and Further Research Questions

If this study has confirmed my beliefs about pluralism in interpretation, it has left me even more skeptical about omnibus definitions of data use which have heretofore prevailed in educational administration studies. For those advocating Knowledge Transfer, the implication is that we must more closely look at the actual interpretive processes and purposes of the reader with particular statistics rather than confusing data with knowledge and assuming some critical mass of high quality information will eventually be attained as a “tipping point” for levering change in schools. More assessment and more data will not automatically lead to improvement (Shen & Cooley, 2008). If Peirce’s theory does not entirely undermine a Modal Technologies perspective which depicts the reader as puppet of the author, Peirce’s propositions certainly demonstrate that the school principal has agency and is accomplishing much more than statecraft when drawing inferences from statistics. Researchers must become much more attuned to educators’ stagecraft when they are constructing classroom marks. In Peirce’s semiotic, the relations between Sign, object and interpretant are diagrammatically and dialectically contrary to Foucault’s panoptic. Social Constructivists may need to broaden their horizons beyond small-group discussion as an all-purpose technique for creating an organizational community. School principals have many other managerial methods at their disposal to moderate or amplify the meaning of a statistic. These methods range
from the conscious use of gesture, to the judicious choice of space, to a verbal and emotional register spanning the interrogative, declarative, and imperative moods. Striking a committee or forming a professional learning circle is a partial but narrow approach to educational improvement. And for those adopting the Enlightenment perspective, the major message is that interest and ideology do interact with information, but at a much more pointillist, micropolitical level than in the broad brush stroke portraits that proponents of enlightened data use suggest.

Across all four orientations to data use, different approaches to conducting research may be required. If anything, survey researchers may both underestimate and underappreciate the impact of statistics because their methods do not “capture” the fleeting traces of gesture, spatial use, and tone of question that principals may pose with various audiences after reading a report. That may be personally reassuring after two decades of work to ensure that educational statistics do not collect dust on a shelf, but it does open expansive vistas for further research. At a minimum, looking at statistical interpretation requires that researchers take into account the varied purposes and nuanced responses that readers have for particular notations, and especially to consider the role of questions in stimulating, guiding or thwarting data use. Given that survey methods will likely prevail when investigating statistical use rather than the qualitative approaches favoured in faculties of education, survey questions should be framed in the conditional future tense. Questions might focus more on school principals’ intentions and their projected effects for various types of statistics and less on simply asking the educator to choose from a predefined menu of Likert scale attitudes that do not distinguish between various types of statistics, or the range of motivations at play.

Nevertheless, Peirce’s theory remains complex in its semiotic ideas, and ethereal in many ontological assumptions. Pragmatism is sometimes mischaracterized as devoid
of an overarching set of ideas or ideals. So too is it sometimes seen as ethically void, as morally sanctioning individual interests, and as underwriting any set of purposes provided the desired end is attained. In other words, classical pragmatism is sometimes mischaracterized as the ends justifying any expedient means, in multiple senses of the word. School principals’ elastic and situational notions of the centre might appear to condone such a view. But that is a definition to which Peirce strenuously objected (EP2. 242-257), insisting that the pragmatic maxim as a method of reflection is guided by a sense of proper or moral conduct that holds community values in consensus as the desired end.

Instead, he set out two deontologically ethical maxims: “Dismiss make-believes” (EP2. 335) and “Do not block the road of inquiry” (EP2. 48). We should resist attempts by others to exercise power by forcing their beliefs or ideals on us, and we should always enable others to pursue their own questions. Whether those injunctions are universal and whether pragmatism’s moral suppleness is appealing will hinge on one’s ethical stance. However, these maxims do provide strong guidance for managers of statistical meaning—administrators will only with difficulty impose their own meaning for an average on others. The primary function of a statistic should be to reflect on autonomous and eventually group practice as a form of professional development, rather than to drive unilateral decision making.

Classical pragmatism’s other shortcomings become clear in research application. Peirce’s brand of phenomenology remains under-articulated and under-employed as a method. Peirce was not interested in the mundane issues of applying his semiotic, preferring to philosophically map the emergent sciences rather than deal with the difficulties of operationalizing his ideas in research. Although William James (1890) had much more to say about attention, perhaps because of his brother’s literary interest in
streams of consciousness and interior monologue, Peirce inadequately described “attention” as an effort of will having only denotative and not connotative functions (EP2. 22-24). This is a serious shortcoming for perception studies, and for any educational research informed by pragmatism. Disentangling cause-effect from antecedent-consequent and distinguishing Peirce’s three kinds of reasoning processes remains a challenge in research application. Researchers have recovered Peirce’s formula for a reader to probabilistically calculate the success of predictions with two alternative answers to a research question, but not his formula for questions which have more than two alternative answers (Calster, 2012; Loken, 2010; Peirce, 1884).

Peirce himself provided perhaps the most astringent commentary on his expansive theory of interpretation: “the maxim of Pragmatism does not bestow a single smile upon beauty, upon moral virtue, or upon abstract truth;—the three things that alone raise Humanity above Animality” (EP2. 465). School principals’ conceptions of average student achievement do not have the elegance of a mathematician’s deductive symmetry, do not offer definitive and unimpeachable statements of goodness and badness as intrinsic virtues, and sometimes demonstrate the brute interests of those who exercise leadership rather than pursuing an abstract truth. High ideals divorced from experience, and all ideologies which a priori codify (in)appropriate conduct, receive short shrift in classical pragmatism. Pragmatists will insist that axiology (the study of values as statements of desire with motivating force) should not be confused with morality (consideration of goodness and badness), nor should morality be confused with ethics (as the study of right and wrong conduct). Peirce admitted that he remained “a determined foe of no innocent number;” the numerals were “thin skeletons of thought” (CP1. 355), by which he intended that statistics are often laden with diverse motivations, hence
moralties, and thereafter stultifying ideologies.²² Moreover, Peirce in his personal life, if not entirely so in his writings, remained hopelessly naïve about the power of capital, unlike William James. The mass media and elites have massive resources available to manufacture statistical evidence. Even if Peirce remained convinced that dissent and an inquiring frame of mind would eventually prevail over any a priori or authoritative method of fixating belief, many will remain skeptical.

Notwithstanding these shortcomings, we may draw on Peirce’s pragmatism itself for three overarching criteria (and thereby turn his thinking in on itself) when evaluating his theory for the interpretation of statistics within educational administration. These criteria are: verisimilitude (conformity of a perception or observation with experience); probability (likelihood of a belief compared with objective chance); and plausibility (conformity of observations in their breadth, economy and reasonableness, independent of any other evidence other than our instincts)(Beeson, 2008; CP8. 222-223; EP2. 441).

Saskatchewan school principals’ interpretations of data displays were verisimilar to many of Peirce’s ideas: (a) predominately abductive reasoning within school principals’ interpretative practices, with an acknowledgement of fallibilism; (b) an awareness of the parallax in others’ interpretations for an identical Sign, and especially the assignment of different objects for that Sign; (c) the recurrent search for (dis)confirming evidence; (d) a rejection of utopian ideas in student assessment; (e) recognition of multiple continua in student learning; (f) trivalent conceptions for the statistic in its iconic, indexical, and symbolic properties; (g) various admixtures of administrative values for the ordinal, ²² Peirce (EP2. 58-61) provided a comprehensive classification of human motives as moralities in ethics when rejecting Karl Pearson’s claim in his Grammar of Science that statistics should serve “to strengthen social stability.” Among the diverse moralities Peirce identified were individualism, utilitarianism, altruism, sentimentalism, historicism, evolutionism, rationalism, dialecticism, educationalism, pancratism, and religionism. There is little evidence in Peirce’s writing that he had read or was interested in political philosophy.
cardinal and nominal properties of numbers; (h) the anti-Cartesian expression of norms and averages through gesture with questioning that has an underlying spatial or locational sense; and (i) the longitudinal evolution of an average conception (Watson & Moritz, 1999, 2000). Moreover, the inventory administered during the second phase of this study revealed (j) small but probabilistically significant changes in school administrators’ self-reported beliefs; (k) probabilistically significant interactions between valence and audience when the average representation is presented as a reading task for interpretation, leading to significant changes in leadership disposition; and (l) a high likelihood that increased self-efficacy, as an element of reader agency, will predict cognitive styles as beliefs. Taken together, these findings present a plausible picture of the school administrator as pragmatic inquirer, if not politician.

“Whatever else it produces,” Kahneman (2011) has declared, “an organization is a factory that manufactures judgements and decisions” (p. 418). In Canadian schools, thousands of such professional judgements are routinely made during a school year by teachers with direction from school principals—when appraising student performances, when constructing assignments and marking student work, and when preparing reports for multiple audiences. To manage the meaning of these statistics, school administrators consider average student achievement not with the inferential patterns assumed within contemporary cognitive science’s notions of heuristic irrationality, but rather as a reasoned form of inquisitive thinking and behaviour which has been formalized and comprehensively described in North American philosophy for over 100 years. To adequately understand the meaning of the statistical average, we must avoid succumbing to what William James (1890) called the “great snare” of the psychologist’s fallacy: “the confusion of his own standpoint with that of the mental fact about which he is making his report” (p. 290)—superimposing our own categories on those of others. Many so-called
Cognitive illusions are better understood as pragmatic, adaptive responses in belief to situations of uncertainty (Gigerenzer, 2008).

For in their dependence on hunches using abductive reasoning with averages, no less than in their proclivity to act upon hunches in a pragmatic fashion, Saskatchewan school principals’ interpretations of average student achievement appear not as extensions of, but as reactions against the excessive rationalism of applying strictly mathematical approaches within a compulsory education system. If we must place these Saskatchewan school principals within an instructional leadership framework—setting aside for the moment Peirce’s constellation of ideas—then school principals’ beliefs are antithetical to the Weberian (Gerrt & Wright Mills, 2009) rationalities of a mass bureaucratization of schooling, and rejections of the convoluted and contradictory logic of significance testing with the null hypothesis (Gliner, Leech & Morgan, 2002; Morgan, 2003). Instead, Saskatchewan school administrators exemplify much older pedagogic traditions, found in Socratic questioning and in Aristotelian mimesis, on one hand, and in Archimedes’ practical experiments for finding the centre of gravity, on the other. It is to those traditions that we must pragmatically turn if we are to better prepare school leaders for helping them conceive of student achievement and for using data for educational improvement.

In light of the foregoing, then, the pertinent research questions (in no particular order) for instructional leadership and for the management of meaning are:

1. What are the implications of Ministry of Education officials asking questions rather than providing authoritative commentary when preparing student achievement reports?
2. (How and why) do government officials reason differently than educators with performance statistics?
3. What are the similarities and differences in adult reading processes for prose and for statistics?
• What are the interrelationships between school administrators’ gestures, sense of space, use of statistics, and questions before various audiences?

• How do student averages that have been disaggregated by gender, ethnicity, and socioeconomic status interact with school administrators’ values?

• What are school administrators’ conceptions of probability?

• What are school administrators’ conceptions for the statistical ideas of variation, distribution, and power?

• How do school principals conceive and calculate risk?

• How do school principals’ proportionally reason with ratio concepts?

• How do principals’ perceptions of averages as representations of quality interact with their conceptions of averages as quantities in memory?

• How do readers’ statistical understandings interrelate with perceptions of a(n) (in)visible author?

• What are First Nations’ traditional conceptions of centre and of student achievement, both within and outside of counting practices?

• What was C. S. Peirce’s mathematical formula for calculating practical significance with research questions having more than two hypothetical answers?

• What are teachers’ and parents’ perceptions and perspectives on both average student achievement and school principals’ questions?
References


Deutsch, M. (1975). Equity, equality, and need: What determines which value will be used as the basis of distributive justice?. *Journal of Social issues, 31*(3), 137-149.


Hallinger, P., & Heck, R. (2002). What do you call people with visions? The role of vision, mission and goals in school leadership and improvement. In K. Leithwood & P. Hallinger (Eds.), The second international handbook of research in educational leadership and administration, Volume 2 (pp. 9-40). Dordrecht, NL: Kluwer.


Irwin P., Levin, I., Schneider, S., & Gaeth, G. (1998). All frames are not created equal: A typology and critical analysis of framing effects. *Organizational Behavior and Human Decision Processes, 76*(2), 149-188.


Levin, I., Schneider, S., & Gaeth, G. (1998). All frames are not created equal: A typology and critical analysis of framing effects. Organizational Behavior and Human Decision Processes 76, 149-188.


357


New directions in judgment under uncertainty. (pp. 582-599). New York: Cambridge University Press.


Appendices

Appendix A: Permissions from the University of Regina and University of Saskatchewan

Appendix B: Informed Consent Forms

Appendix C: Permission for Use of Rational Experiential Inventory

Appendix D: Table of Correspondence for Research Question #1, Conjectures by Semi-Structured Interview Question Numbers in Phase 1

Appendix E: Phase 1: Semi-Structured Interview Questions

Appendix F: Phase 1: Interview Data Displays

Appendix G: Phase 2: Four Frames for Interpreting the Average

Appendix H: Epstein’s Dual Process Model

Appendix I: Matrix of Findings Cross tabulated with Data Source

Appendix J: Recommendations for Policy, Program and Practice
Appendix A: Permissions from the University of Regina and University of Saskatchewan

UNIVERSITY OF REGINA

OFFICE FOR RESEARCH, INNOVATION AND PARTNERSHIP
MEMORANDUM

DATE: November 15, 2012

TO: Darryl Hunter
1231 Rossie Drive
Regina, SK S4N 6Z6

FROM: Dr. Larena Hoeber
Chair, Research Ethics Board

Re: About Average: A Pragmatic Inquiry into School Administrator’s Meanings for a Statistical Concept in Instructional Leadership (File # 28S1213)

Please be advised that the University of Regina Research Ethics Board has reviewed your proposal and found it to be:

☐ 1. APPROVED AS SUBMITTED. Only applicants with this designation have ethical approval to proceed with their research as described in their applications. For research lasting more than one year (Section 1F). ETHICAL APPROVAL MUST BE RENEWED BY SUBMITTING A BRIEF STATUS REPORT EVERY TWELVE MONTHS. Approval will be revoked unless a satisfactory status report is received. Any substantive changes in methodology or instrumentation must also be approved prior to their implementation.

☐ 2. ACCEPTABLE SUBJECT TO MINOR CHANGES AND PRECAUTIONS (SEE ATTACHED). Changes must be submitted to the REB and approved prior to beginning research. Please submit a supplementary memo addressing the concerns to the Chair of the REB.** Do not submit a new application. Once changes are deemed acceptable, ethical approval will be granted.

☐ 3. ACCEPTABLE SUBJECT TO CHANGES AND PRECAUTIONS (SEE ATTACHED). Changes must be submitted to the REB and approved prior to beginning research. Please submit a supplementary memo addressing the concerns to the Chair of the REB.** Do not submit a new application. Once changes are deemed acceptable, ethical approval will be granted.

☐ 4. UNACCEPTABLE AS SUBMITTED. The proposal requires substantial additions or redesign. Please contact the Chair of the REB for advice on how the project proposal might be revised.

Dr. Larena Hoeber

cc: Dr. Rod Dolmage - Education

** supplementary memo should be forwarded to the Chair of the Research Ethics Board at the Office for Research, Innovation and Partnership (Research and Innovation Centre, Room 109) or by e-mail to research.ethics@uregina.ca

Phone: (306) 585-4775
Fax: (306) 585-4893

363
Dr. Rod Dolmage  
March 8, 2013  
RE: About Average: A Pragmatic Inquiry into School Administrator’s Meanings for a Statistical Concept in Instructional Leadership  

Student: Daryl Hunter  
U of R File #: 28S1213  
U of S File #: BEH 12-311  

Your application for research ethics approval has undergone a harmonized review by the University of Regina and the University of Saskatchewan. The University of Saskatchewan REB acknowledges that it has had the opportunity to participate in the review of your application. A Certificate of Approval has been issued by the University of Regina.  

In accordance with the Research Ethics Review Reciprocity Agreement signed by the University of Saskatchewan, University of Regina, and Regina Qu’Appelle Health Region dated June 1, 2012, the University of Saskatchewan REB accepts the Certificate of Approval issued by the REB of the University of Regina. This letter acknowledging acceptance of a reciprocal research ethics review is issued to you in lieu of a Certificate of Approval by the University of Saskatchewan REB. This letter permits you to conduct research activities as approved by University of Regina REB, provided that you maintain a valid and up-to-date Certificate of Approval.  

All continuing ethics review will be conducted by the University of Regina REB. The University of Regina is authorized to share all communications pertaining to this file with the University of Saskatchewan REB at their discretion. The University of Saskatchewan REB may provide input into continuing ethical review activities, as agreed upon by both REBs.  

The University of Saskatchewan REB reserves the right to revoke the privileges described in this letter at any time in order to conduct their own independent research ethics review of your project. Such a decision would be communicated to you and the University of Regina REB in writing.  

Best wishes for your continuing research endeavours.  

Sincerely,  

Beth Bilson, Chair  
Behavioural Research Ethics Board  
University of Saskatchewan
Appendix B: Informed Consent Forms

CONSENT FORM (Phase 1)

Name of Research Project: “About average: A pragmatic study of school administrators’ conceptions of a statistical concept for instructional leadership”

This study of school administrator’s representations of the statistical average is being conducted by Darryl Hunter for the University of Regina, Faculty of Graduate Studies and Research, in partial fulfillment of the requirements for a PhD program.

Purposes – The purpose of the study is to better understand school administrators’ concepts of an average in relation to student achievement. The study aims to further our notions of assessment literacy, and better prepare school leaders for roles involving instructional leadership. The goal is to find better ways of presenting statistics to make it easier for school-leaders to interpret information for targeted problem-solving and decision-making.

Procedure – The objective in this first phase of the study is to gain a better understanding of how/what you think about averages. Three rounds of interviews are forecast (total 3 hrs.): the first is to understand your notion of an average; the second is to understand how you would apply that concept in different situations, and; the third will look specifically at your interpretation of the statistic as you read a statistical chart.

Each interview will be tape-recorded. I will also ask for diagrams, pictures or other images of your averages. I forecast using the material for a dissertation, for conference proceedings and for scholarly publication.

Confidentiality – Your participation will remain confidential. General approval to conduct this study has been received from the school board office. However, your name was suggested as a potential respondent among several nominees, so no link can be drawn between you and your commentary. Pseudonyms for you, your school and your school division will be used in all publications.

Interview data (tapes, interview transcripts and associated materials) will be stored in a locked filing cabinet at the University of Regina, Faculty of Education Rm. 235.

Permissions – The study has been reviewed and approved by the Research Ethics Board at the University of Regina. Feel free to contact the Research Ethics Board if you have any questions about your rights and treatment as a participant at (306) 585-4775 or research.ethics@uregina.ca.

I will share transcripts of our tape-recorded interviews with you, so that you may correct them for accuracy and clarify any matter. I will also share a first draft of relevant dissertation material that I write, so you can confirm my overall interpretation.

Withdrawal – Should you have questions about study procedures and findings, want to withdraw your data, or access your transcripts, please contact either one of us:

- Darryl Hunter, Regina at hunter5d@uregina.ca
- Dr. Rod Dolmage, Rm. 236.2, Faculty of Education, Regina (306) 585-4816 or rod.dolmage@uregina.ca

Data will be analyzed and substantial research writing will be accomplished in the January to March 31, 2013 interval. Data cannot be withdrawn after March 31, 2013.

Consent – Please signify your consent to participate by signing and dating the line below. You are free to withdraw from the study at any point, or to not participate at all. Participation is voluntary. At the same time, you may ask for me to return images or materials that you generate in the context of this study.

Signature of Respondent/Date

Signature of Researcher/Date
(You will receive a copy of the signed consent form)
Welcome to this unique research opportunity!

**Name of Research Project:** “About average: A pragmatic study of school administrators’ conceptions of a statistical concept for instructional leadership”

**Purposes**
The purpose is to study school administrators’ decision-making styles in relation to the statistical concept of the average. The study will inform the preparation of instructional leaders, contribute to our notions of “assessment literacy” in education circles, and suggest better ways of fostering numeracy among educators.

**Contacts**
This study of school administrators’ understanding of the statistical average is being conducted by Darryl Hunter for the University of Regina, Faculty of Graduate Studies and Research, in partial fulfillment of the requirements for a PhD program. His contact information is:

- Darryl Hunter, Regina at hunter5d@uregina.ca
- Dr. Rod Dolmage, Rm. 236.2, Faculty of Education, Regina (306) 585-4816 or rod.dolmage@uregina.ca.

**Procedure**
The study proceeds in three steps. You will

1. complete a questionnaire which asks for basic information about your background, and about your decision-making preferences
2. read and interpret graphic material, statistics and commentary on your own
3. complete a follow-up questionnaire which asks about your decision-making preferences

All three steps will be completed today. This entirely online study should take approximately 20 to 30 minutes to complete.

**Anonymity**
Within the limits possible in a group (meeting) context, your participation and individual responses will remain confidential. Only a complete, cumulative profile without any personal information will be stored. Your anonymity is preserved in all reporting. All identifiable codes will be stripped from the data files after March 31, 2013. That data will be stored at the University of Regina and at the University of Saskatchewan in password protected files which are encrypted, locked and protected by university electronic and facility security protocols.

Only the researcher conducting this study will have access to the complete data set. Once the study is completed, only aggregate and summary profile data will be reported. The data will be used to develop papers for publication in academic journals and conference proceedings.

For your protection, this research project has been approved by the University of Regina’s and University of Saskatchewan’s Research Ethics Boards. Any questions regarding your rights as a participant may be addressed to either of those Boards through the Ethics Office at (306) 585-4775 (Regina) or (306) 966-2084 (Saskatoon). Out of town participants may call collect.

**Consent**
Your participation in this study is voluntary. You may withdraw from the study for any reason and at any time without penalty. Your right to withdraw data from the study will apply until the data have been pooled on or about March 31, 2013. After this, it is possible that some form of research dissemination will have already occurred and it will not be possible to withdraw your data.

Completion of the study signifies that you have understood the above information regarding your participation in this project and that you have agreed to participate. In no way does this release the researcher or involved institutions from their professional responsibilities to abide by national ethical standards.

Thank you in advance for your participation in this project. Your contribution is appreciated.
Appendix C: Permission for Use of Rational Experiential Inventory

Dear Daryl,

With this e-mail I grant you permission to use the REI for research purposes. Attached at the REI, its scoring key, and a summary of the research we have done with it. If you use it in your research, I would appreciate a summary of your findings.

Best wishes,

Seymour Epstein
Appendix D: Table of Correspondence for Research Question #1, Conjectures by Semi-structured Interview Question Numbers in Phase 1

<table>
<thead>
<tr>
<th>Research Question</th>
<th>Conjectures</th>
<th>Interview 1</th>
<th>Interview 2</th>
<th>Interview 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 a)</td>
<td>Conceptions and calculations of an average that are nuanced and multiple</td>
<td>2. 4. 5. 6.</td>
<td>1. 2. 7.</td>
<td>7. 8.</td>
</tr>
<tr>
<td>1 b)</td>
<td>Inferences from statistical representations in relation to curriculum priorities and student performances</td>
<td>3. 7.</td>
<td>3. 5.</td>
<td>2. 4.</td>
</tr>
<tr>
<td>1 c)</td>
<td>Consequences and effects in relation primarily to classroom rather than large-scale assessment results</td>
<td>7. 8. 13.</td>
<td>3. 5.</td>
<td>1. 2. 6.</td>
</tr>
<tr>
<td>1 d)</td>
<td>Position as causal agents with statistical representations when carrying out their administrative roles</td>
<td>1. 2. 9.</td>
<td>4. 5. 6.</td>
<td>2. 3. 4.</td>
</tr>
<tr>
<td></td>
<td></td>
<td>11. 12.</td>
<td>8. 10. 11.</td>
<td>5. 7. 8.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>12. 15. 16.</td>
<td></td>
</tr>
</tbody>
</table>

*Note:* Numbers under each interview refer to semi-structured question numbers in Appendix E interview guides.
Appendix E: Phase 1: Semi-structured Interview Questions

Semi-Structured Interview Guide: Interview 1

Purpose of Interview: General practical sense of average in relation to leadership purpose
(Prompt questions in brackets)

1. Could you tell me about your background and education? How and why did you choose to become a school principal?
2. So would you say that your school is average, above average, below average? (Why do you say that? How do you know?)
3. In your view, what/who would be an average student? (as an individual, as a class, as a subgroup?) (Is there a difference between a normal, typical, average student?)
4. What does an average mean to you? What are the major characteristics of an average?
5. Can you describe some experiences which were memorable in terms of “averages”?
6. What are the differences between “being average” and “knowing the average”?
7. What are the short-term and long-term consequences of being average?
8. What experience(s) do you have with statistics? (re statistics hard or soft? Why do you say so?)
9. What are the purposes for calculating an average? Why do you (not) calculate an average?
10. (How) are average student attendance, average class size, average per pupil expenditure, average student achievement and average marks different? (How would you compute the average in the same way or differently in each case?)
11. What role do statistics play (not play) in your leadership practice?
12. How do you see your role in relation to student achievement?
13. How would an above/at/below average statement in outcomes have consequences for you in your leadership?
14. What are the elements or constituents in “your” average?
15. What questions should I be asking that I am not?
Semi-structured Interview Guide: Interview 2

Purpose of Interview: Computational act and detail in terms of consequences.

1. How would you now compute an average for individual student performance as school principal? How would you compute an average for entire school performance?
2. What images for pictures come to mind when you see an average? Could you sketch them for me? If you were to make a graph or statistical picture of the word “average”, what would it look like? Could you show me?
3. What consequences or effects do averages have for students? parents? families? community? central office?
4. What does an average mean to you? What are the major characteristics of an average?
5. If you were asked to diagram average student achievement and the relationships between the student and its consequences, how would you? Where do you fit in that diagram?
6. If you were asked to work with someone to interpret statistical averages, who would you choose? With other principals?
7. If you were asked to compute an average student achievement, what/who would be included and excluded from your average?
8. If you have a conflict with parents over averages, how do you handle it?
9. How would you handle students who leave-arrive at the school in terms of who/how many are included in the denominator?
10. How do you conceive of a typical school? How do you conceive of an atypical school?
11. How do you explain (away) below average marks to parents? to the employer? to parents?
12. What is the role of a Learning Improvement Plan? What is role of the Ministry’s Continuous Improvement Framework? What is the role of assessment in determining average student achievement?
13. What aspects of school life can’t be averaged? Why not?
14. What would be an intolerable consequence from a statistically below average? How would you act?
15. Statistical averages are to decisions as x is to y?
16. If Saskatchewan or Canada had below average student achievement, I would …
Semi-structured Interview Guide: Interview 3

**Purposes:** Probe for greater detail on comments/script from Interview 1 and 2 in terms of consequences. Position of the principal in terms of computations, consequences, effects and decision making.

1. What is the difference between a statistic having an effect, having consequences, and causing something to happen? or someone to change?
2. What does a statistical average mean in terms of your interactions with staff, school, students, parents?
3. How would you communicate an average to the School Community Council?
4. What role do averages have for your school action plans?
5. What is the difference between problem-solving and decision-making?
6. Would you say that statistical averages create or help solve problems? Why? What would be a better way of creating or solving problems?
7. What challenges do you have communicating above, below, at average student achievement?
8. How could central office help you in computing and calculating the statistical average?
## Appendix F: Phase 1: Interview Data Displays

### Data Display 1: Elementary

<table>
<thead>
<tr>
<th>English Language Arts</th>
<th>Maths</th>
<th>Science</th>
<th>Health</th>
<th>Arts</th>
<th>Attendance</th>
<th>Provincial Assessment in Reading</th>
<th>Division-wide Canadian Standardized Reading Level Grade Equivalent</th>
<th>Division-wide Canadian Standardized Math Level Grade Equivalent</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anson, Joan</td>
<td>2</td>
<td>40</td>
<td>65</td>
<td>2</td>
<td>12</td>
<td>65 absent</td>
<td>Grade 5</td>
<td>Grade 2</td>
<td>Mother is school board chair</td>
</tr>
<tr>
<td>Bieber, Jeffery</td>
<td>2</td>
<td>47</td>
<td>65</td>
<td>3</td>
<td>18</td>
<td>87</td>
<td>Grade 6</td>
<td>Grade 1</td>
<td>Mother is mayor</td>
</tr>
<tr>
<td>Johnson, Reena</td>
<td>4</td>
<td>56</td>
<td>66</td>
<td>3</td>
<td>9</td>
<td>85</td>
<td>Grade 6</td>
<td>Grade 3</td>
<td>Excellent athlete</td>
</tr>
<tr>
<td>Manton, Mary</td>
<td>3</td>
<td>49</td>
<td>65</td>
<td>1</td>
<td>17</td>
<td>87</td>
<td>Grade 6</td>
<td>Grade 3</td>
<td></td>
</tr>
<tr>
<td>Pradhue, Hassan</td>
<td>2</td>
<td>50</td>
<td>65</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>4 not registered at time of test</td>
<td>Grade 3</td>
<td>Neo-immigrant; English as Second Language Student</td>
</tr>
<tr>
<td>Rockwood, Allen</td>
<td>4</td>
<td>68</td>
<td>65</td>
<td>4</td>
<td>9</td>
<td>80</td>
<td>3 Grade 7</td>
<td>Grade 5</td>
<td></td>
</tr>
<tr>
<td>Stonechild, Judith</td>
<td>3</td>
<td>32</td>
<td>0</td>
<td>4</td>
<td>18</td>
<td>55</td>
<td>4 Absent</td>
<td>Grade 2</td>
<td>On reserve; attended FN school last year</td>
</tr>
<tr>
<td>Truman, Henry</td>
<td>4</td>
<td>67</td>
<td>63</td>
<td>4</td>
<td>12</td>
<td>86</td>
<td>3 Grade 6</td>
<td>Grade 4</td>
<td></td>
</tr>
<tr>
<td>Velts, Mark</td>
<td>1</td>
<td>77</td>
<td>65</td>
<td>4</td>
<td>16</td>
<td>87</td>
<td>4 Grade 5</td>
<td>Grade 5</td>
<td></td>
</tr>
<tr>
<td>Whiteman, Jack</td>
<td>4</td>
<td>55</td>
<td>64</td>
<td>4</td>
<td>16</td>
<td>87</td>
<td>4 Grade 7</td>
<td>Grade 6</td>
<td></td>
</tr>
</tbody>
</table>

X/4 point scale where 4 is high

X/100

X/100

X/18 number of curriculum outcomes student has minimally demonstrated capability in

X/87 classes

X/4 point scale where 4 is high

Reading grade level

Math problem-solving level

### Data Display 1: Secondary

<table>
<thead>
<tr>
<th>English 10</th>
<th>Math 10</th>
<th>Science 10</th>
<th>Wellness 10</th>
<th>Arts Education 10</th>
<th>Attendance</th>
<th>Provincial Assessment</th>
<th>Division-wide test: Canadian Standardized Reading Level</th>
<th>Division-wide test: Canadian Standardized Math Level</th>
<th>Notes</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anderson, James</td>
<td>81</td>
<td>48</td>
<td>65</td>
<td>85</td>
<td>56</td>
<td>65 absent</td>
<td>Grade 9</td>
<td>Grade 5</td>
<td>Mother is school board chair</td>
</tr>
<tr>
<td>Amil, Tamir</td>
<td>51</td>
<td>50</td>
<td>65</td>
<td>95</td>
<td>44</td>
<td>5 not registered at time of test</td>
<td>Grade 2</td>
<td>Grade 8</td>
<td>Neo-immigrant ESL student</td>
</tr>
<tr>
<td>Bohach, Mary</td>
<td>47</td>
<td>47</td>
<td>65</td>
<td>88</td>
<td>59</td>
<td>87</td>
<td>3 Grade 6</td>
<td>Grade 10</td>
<td>Mother is mayor</td>
</tr>
<tr>
<td>Johnson, Richard</td>
<td>99</td>
<td>56</td>
<td>66</td>
<td>81</td>
<td>33</td>
<td>85</td>
<td>3 Grade 11</td>
<td>Grade 4</td>
<td>Local Junior B hockey star</td>
</tr>
<tr>
<td>Kennedy, Helen</td>
<td>88</td>
<td>67</td>
<td>63</td>
<td>96</td>
<td>61</td>
<td>86</td>
<td>3 Grade 10</td>
<td>Grade 9</td>
<td>Father is Superintendent</td>
</tr>
<tr>
<td>Mazokowski, Alicia</td>
<td>79</td>
<td>49</td>
<td>65</td>
<td>96</td>
<td>60</td>
<td>87</td>
<td>3 Grade 10</td>
<td>Grade 11</td>
<td></td>
</tr>
<tr>
<td>Redman, Joan</td>
<td>96</td>
<td>55</td>
<td>64</td>
<td>89</td>
<td>66</td>
<td>87</td>
<td>4 Grade 12</td>
<td>Grade 9</td>
<td></td>
</tr>
<tr>
<td>Rockford, Allison</td>
<td>89</td>
<td>68</td>
<td>65</td>
<td>87</td>
<td>60</td>
<td>80</td>
<td>3 Grade 7</td>
<td>Grade 9</td>
<td></td>
</tr>
<tr>
<td>Stonechild, Ralph</td>
<td>65</td>
<td>32</td>
<td>0</td>
<td>98</td>
<td>87</td>
<td>55</td>
<td>4 Absent</td>
<td>Grade 10</td>
<td>On reserve; attended FN school last year</td>
</tr>
<tr>
<td>Vonish, Mary</td>
<td>33</td>
<td>77</td>
<td>65</td>
<td>88</td>
<td>59</td>
<td>87</td>
<td>4 Grade 10</td>
<td>Grade 10</td>
<td></td>
</tr>
</tbody>
</table>

X/100

X/100

X/100

X/87 classes

X/4 point scale where 4 is high

Reading grade level

Math problem-solving level equivalent
Imagine that your school has received test outcomes, where average student results include a provincial comparison.

You have been asked to prepare and lead an action plan to improve student results. The anticipated consequences of your plan are as follows:

- If your plan is adopted, 80% of the students in your school will demonstrate below average or average results next year*.

- Both your plan and results will be made public.

* Statistically different than the provincial average, \( p < .001 \).
SCENARIO 3

SCENARIO: I’m going to show you a data display of the kind you might have if you are comparing your school to your district. This is a bar graph of Grade 3 reading achievement separated into two components (fluency and comprehension) as well as their total score, for Lake Forest School and its district for each of 3 years. (Show Figure 2.)

Figure 2. Grade 3 Reading Achievement Scores Over Three Years
Figure 12a: Average Marks for Selected 10 and 20 Level Courses, by Student Characteristic, 2009-10

<table>
<thead>
<tr>
<th></th>
<th>Non-Aboriginal</th>
<th>Urban Male</th>
<th>Urban Female</th>
<th>Rural Male</th>
<th>Rural Female</th>
<th>North Male</th>
<th>North Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Language Arts A 10</td>
<td>68.4%</td>
<td>76.2%</td>
<td>70.0%</td>
<td>78.2%</td>
<td>65.7%</td>
<td>70.1%</td>
<td></td>
</tr>
<tr>
<td>English Language Arts B 10</td>
<td>68.1%</td>
<td>75.9%</td>
<td>70.5%</td>
<td>78.9%</td>
<td>61.3%</td>
<td>63.9%</td>
<td></td>
</tr>
<tr>
<td>Science 10</td>
<td>69.1%</td>
<td>73.5%</td>
<td>69.8%</td>
<td>75.5%</td>
<td>62.2%</td>
<td>64.9%</td>
<td></td>
</tr>
<tr>
<td>Mathematics 10</td>
<td>68.6%</td>
<td>73.3%</td>
<td>71.1%</td>
<td>76.5%</td>
<td>58.5%</td>
<td>59.1%</td>
<td></td>
</tr>
<tr>
<td>English Language Arts 20</td>
<td>67.8%</td>
<td>76.5%</td>
<td>70.3%</td>
<td>80.6%</td>
<td>57.9%</td>
<td>68.0%</td>
<td></td>
</tr>
<tr>
<td>Biology 20</td>
<td>67.6%</td>
<td>73.6%</td>
<td>70.0%</td>
<td>77.9%</td>
<td>62.3%</td>
<td>73.4%</td>
<td></td>
</tr>
<tr>
<td>Mathematics 20</td>
<td>67.7%</td>
<td>71.9%</td>
<td>69.8%</td>
<td>75.4%</td>
<td>63.9%</td>
<td>66.5%</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Self-Declared Aboriginal</th>
<th>Urban Male</th>
<th>Urban Female</th>
<th>Rural Male</th>
<th>Rural Female</th>
<th>North Male</th>
<th>North Female</th>
</tr>
</thead>
<tbody>
<tr>
<td>English Language Arts A 10</td>
<td>55.7%</td>
<td>62.3%</td>
<td>55.2%</td>
<td>61.1%</td>
<td>49.8%</td>
<td>58.6%</td>
<td></td>
</tr>
<tr>
<td>English Language Arts B 10</td>
<td>56.6%</td>
<td>61.8%</td>
<td>53.4%</td>
<td>60.6%</td>
<td>50.6%</td>
<td>58.1%</td>
<td></td>
</tr>
<tr>
<td>Science 10</td>
<td>53.0%</td>
<td>55.3%</td>
<td>54.0%</td>
<td>59.2%</td>
<td>51.0%</td>
<td>55.1%</td>
<td></td>
</tr>
<tr>
<td>Mathematics 10</td>
<td>54.0%</td>
<td>56.8%</td>
<td>52.6%</td>
<td>57.1%</td>
<td>49.7%</td>
<td>54.5%</td>
<td></td>
</tr>
<tr>
<td>English Language Arts 20</td>
<td>57.1%</td>
<td>62.9%</td>
<td>55.6%</td>
<td>63.4%</td>
<td>57.3%</td>
<td>58.8%</td>
<td></td>
</tr>
<tr>
<td>Biology 20</td>
<td>56.0%</td>
<td>57.4%</td>
<td>58.1%</td>
<td>62.1%</td>
<td>58.7%</td>
<td>61.0%</td>
<td></td>
</tr>
<tr>
<td>Mathematics 20</td>
<td>57.5%</td>
<td>59.2%</td>
<td>56.3%</td>
<td>58.9%</td>
<td>54.7%</td>
<td>57.9%</td>
<td></td>
</tr>
</tbody>
</table>

Figure 12b: Range of Average Marks for Selected 10 and 20 Level Courses, by Student Characteristic, 2009-10

Note: The subgroups shown in Figure 12b had the highest and lowest average marks in the Grade 10 subjects listed. Data includes all schools and all programs, except modified advanced (A, AP, IB), modified basic (M1) courses and courses delivered in French. Registrations plus a mark is required for a mark to be included (excludes transfer credits). Data is categorized as rural or urban based on the predominant residential location of the students served by the school, rather than the location of the school division. Urban enrolments include those students residing in the 15 cities in Saskatchewan; northern enrolments include those residing in the northern region of the province and rural enrolments include all other students.

Appendix G: Phase 2: Four Frames for Interpreting the Average

Frame 1
Imagine that you have test outcomes for your school, where average student results include a provincial comparison. You have been asked to prepare and lead an action plan to improve student results. The anticipated consequences of your plan are as follows:

- If your plan is adopted, 20% of the students in your school will demonstrate above average results next year*.
- Both your plan and results will remain confidential with your teaching staff.

* Statistically different than the provincial average, $p < .001$.

Frame 2
Imagine that you have test outcomes for your school, where average student results include a provincial comparison. You have been asked to prepare and lead an action plan to improve student results. The anticipated consequences of your plan are as follows:

- If your plan is adopted, 80% of the students in your school will demonstrate below average or average results next year*.
- Both your plan and results will remain confidential with your teaching staff.

* Statistically different than the provincial average, $p < .001$.

Frame 3
Imagine that your school has received test outcomes, where average student results include a provincial comparison. You have been asked to prepare and lead an action plan to improve student results. The anticipated consequences of your plan are as follows:

- If your plan is adopted, 20% of the students in your school will demonstrate above average results next year*.
- Both your plan and results will be made public.

* Statistically different than the provincial average, $p < .001$.

Frame 4
Imagine that your school has received test outcomes, where average student results include a provincial comparison. You have been asked to prepare and lead an action plan to improve student results. The anticipated consequences of your plan are as follows:

- If your plan is adopted, 80% of the students in your school will demonstrate below average or average results next year*.
- Both your plan and results will be made public.

* Statistically different than the provincial average, $p < .001$. 

376
Appendix H: Epstein’s Dual Process Model

<table>
<thead>
<tr>
<th>EXPERIENTIAL SYSTEM 1</th>
<th>RATIONAL SYSTEM 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Holistic</td>
<td>• Analytic</td>
</tr>
<tr>
<td>• Emotional: pleasure-pain oriented (what feels good)</td>
<td>• Logical: reason oriented (what is sensible)</td>
</tr>
<tr>
<td>• Associationistic connections</td>
<td>• Cause-and-effect connections</td>
</tr>
<tr>
<td>• More outcome oriented</td>
<td>• More process oriented</td>
</tr>
<tr>
<td>• Behaviour mediated by “vibes” from past experience</td>
<td>• Behaviour mediated by conscious appraisal of events</td>
</tr>
<tr>
<td>• Encodes reality in concrete images, metaphors, and narratives</td>
<td>• Encodes reality in abstract symbols, words, and numbers</td>
</tr>
<tr>
<td>• More rapid processing: oriented toward immediae action</td>
<td>• Slower processing: oriented toward delayed action</td>
</tr>
<tr>
<td>• Slower to change: changes with repetitive or intense experience</td>
<td>• Changes more rapidly: changes with speed of thought</td>
</tr>
<tr>
<td>• More crudely differentiated: broad categorical thinking</td>
<td>• More highly differentiated; dimensional thinking</td>
</tr>
<tr>
<td>• More crudely integrated: dissociative, organized in part by emotional complexes</td>
<td>• More highly integrated</td>
</tr>
<tr>
<td>• Experienced passively and pre-consciously (We are seized by our emotions)</td>
<td>• Experienced actively and consciously: (We are in control of our thoughts)</td>
</tr>
<tr>
<td>• Self-evidently valid: experiencing is believing</td>
<td>• Requires justification via logic and evidence</td>
</tr>
</tbody>
</table>

*Adapted from Epstein, S. (2003).*
Appendix I: Matrix of Findings Cross Tabulated with Data Source

<table>
<thead>
<tr>
<th>Major Finding</th>
<th>I</th>
<th>F</th>
<th>P</th>
<th>D</th>
<th>C</th>
<th>REI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nuanced conceptions for the average in multiple representations, but not mathematically-sophisticated concepts for either measures of central tendency or a norm.</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Average as a centre-of-balance; computational average as an arithmetic mean weighted in relation to curriculum priorities for individual students, mid-range or modal distribution for grade level or classroom marks, and simple arithmetic mean of several key curriculum domains for school level results.</td>
<td></td>
<td>●</td>
<td>●</td>
<td>●</td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Aware of the long-term socio-educational effects of report card averages, and therefore put a priority on classroom marks rather than large-scale assessment results. Largely abductive reasoning for any type of assessment result.</td>
<td>●</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Do not discount notions of significance as the comparison of two means, but most often considered averages across a variety of assessments rather than within a single data set. Averages deriving from budgetary allocations and class-sizes such as PTR deemed more relevant when defining instructional arrangements.</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>School administrators adopted less rational and analytic problem-solving dispositions when school averages were framed negatively than when they were framed positively.</td>
<td></td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>School administrators did not adopt either more rational-analytic or experiential-intuitive problem-solving approaches when school averages were framed as confidential to the school, than when framed as public.</td>
<td>●</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
</tr>
<tr>
<td>Vice-principals had similar cognitive styles to principals.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td></td>
</tr>
<tr>
<td>Experience as a teacher and administrator, position, gender, and education level did not uniquely predict school administrators’ problem-solving dispositions. However, statistical self-efficacy did.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>●</td>
<td></td>
</tr>
</tbody>
</table>

Note: I = interview transcripts, F = field notes, P = photographs, D = diagrams, C = correspondence, REI = Rational Experiential Inventory
### Appendix J: Recommendations for Policy, Program and Practice

<table>
<thead>
<tr>
<th>Recommendation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Provide preservice teachers with experience in critically interpreting (classroom) statistics in terms of sample size, probability, assessment error, and documentary justification</td>
</tr>
<tr>
<td>Model lateral thinking using a variety of attendance and achievement statistics with preservice teachers, emphasizing consequent-antecedent reasoning rather than deductive and inductive inferencing</td>
</tr>
<tr>
<td>Provide preservice teachers with multiple conceptions of the norm and multiple models of data-informed decision-making</td>
</tr>
<tr>
<td>Adopt the Harvard Business School case study approach (Hammond, 1976) when preparing school administrators to interpret statistical information before a variety of audiences</td>
</tr>
<tr>
<td>Provide prospective school administrators with training in Socratic questioning techniques, and in the art of framing statistics with particular audiences and particular intentions in mind</td>
</tr>
<tr>
<td>Provide prospective school administrators with experience in leadership presentation, including the use of gesture, importance of location, and conversational approaches with statistics (Streeck, 2009)</td>
</tr>
<tr>
<td>Provide Saskatchewan educators with information about and experience with multiple models of instructional leadership</td>
</tr>
<tr>
<td>Provide inservice education to educators in the ethics of data presentation, using Tufte’s (1983, 1997) books as reference texts</td>
</tr>
<tr>
<td>Prepare and distribute to educators a handbook on the ethics of statistical interpretation and on conducting crucial conversations with principals, parents, and students with a variety of purposes in mind</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Purpose</th>
</tr>
</thead>
<tbody>
<tr>
<td>To strengthen teachers’ ability to critically interpret and appraise their marks for subsequent improvement of their classroom grading and instructional practices</td>
</tr>
<tr>
<td>To enhance educators’ problem-solving skills relating to student achievement issues</td>
</tr>
<tr>
<td>To strengthen educators’ understanding of normative achievement and applications of evidence-informed choice</td>
</tr>
<tr>
<td>To strengthen school administrators’ ability to apply statistics appropriately in a range of real-life situations</td>
</tr>
<tr>
<td>To enhance the transactional leadership skills of in-school administrators</td>
</tr>
<tr>
<td>To improve the stagecraft of school administrators</td>
</tr>
<tr>
<td>To improve the instructional leadership skills of teachers and in-school administrators</td>
</tr>
<tr>
<td>To enable educators to communicate statistics to various audiences in an ethical manner</td>
</tr>
<tr>
<td>To develop Saskatchewan educators’ ability to question, interpret and use statistics in an appropriate manner.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Responsibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>Undergraduate education programs in Saskatchewan</td>
</tr>
<tr>
<td>Undergraduate education programs in Canada</td>
</tr>
<tr>
<td>Undergraduate education programs in Canada</td>
</tr>
<tr>
<td>Graduate educational leadership programs in Canada</td>
</tr>
<tr>
<td>Graduate educational leadership programs in Canada</td>
</tr>
<tr>
<td>Graduate curriculum, instruction, and educational administration programs in Canada</td>
</tr>
<tr>
<td>Teachers’ federations in Canada</td>
</tr>
<tr>
<td>Saskatchewan Teachers’ Federation</td>
</tr>
<tr>
<td>Action</td>
</tr>
<tr>
<td>----------------------------------------------------------------------</td>
</tr>
<tr>
<td>Strike inter-school interpretation panels comprised of school</td>
</tr>
<tr>
<td>administrators and teachers which periodically review and make</td>
</tr>
<tr>
<td>recommendations for attaining equity by gender, ethnicity and</td>
</tr>
<tr>
<td>geography in classroom assessment marks generated at the student</td>
</tr>
<tr>
<td>and school levels</td>
</tr>
<tr>
<td>Regularly compute, analyze and communicate to teachers and the public</td>
</tr>
<tr>
<td>classroom and large-scale assessment data according to gender, ethnic,</td>
</tr>
<tr>
<td>and geographic categories</td>
</tr>
<tr>
<td>Develop policy for confidentially distributing large-scale assessment</td>
</tr>
<tr>
<td>findings to schools at least 10 days ahead of public release so that</td>
</tr>
<tr>
<td>school principals can meet to develop a consensus on the meaning of</td>
</tr>
<tr>
<td>various averages prior to their becoming public</td>
</tr>
<tr>
<td>Prepare and make available to educators, and to the public,</td>
</tr>
<tr>
<td>foundational documents and detailed technical reports which specify</td>
</tr>
<tr>
<td>purposes and provide formulae, criterion statements, and student</td>
</tr>
<tr>
<td>achievement exemplars for each large-scale assessment</td>
</tr>
<tr>
<td>Provide to teachers and school boards explicit policy guidelines for</td>
</tr>
<tr>
<td>the appropriate weighting of particular outcomes within every</td>
</tr>
<tr>
<td>curriculum guide</td>
</tr>
<tr>
<td>Use the geometric mean when computing (inter)national large-scale</td>
</tr>
<tr>
<td>assessment findings, within or outside applications of item response</td>
</tr>
<tr>
<td>theory</td>
</tr>
<tr>
<td>Revise and update provincial freedom of information legislation to</td>
</tr>
<tr>
<td>emphasize confidentiality of test information at the individual student</td>
</tr>
<tr>
<td>level, and to require that those requesting school or school division</td>
</tr>
<tr>
<td>level information publish that information in a manner consistent with</td>
</tr>
<tr>
<td>the purposes for which it was originally collected</td>
</tr>
<tr>
<td>Revise and update <em>The Freedom of Information Act</em> in Saskatchewan to</td>
</tr>
<tr>
<td>permit and safeguard the electronic transmission of a student’s</td>
</tr>
<tr>
<td>complete cumulative records from Pre-K to undergraduate education</td>
</tr>
</tbody>
</table>