ESTIMATING THE IMPACT OF DISTANCE EDUCATION ON STUDENT LEARNING OUTCOMES USING THE ETS MAJOR FIELD TEST IN BUSINESS

A Dissertation
Submitted to
the Temple University Graduate Board

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

By
Brian Ransom Finnegan
August, 2012

Examining Committee Members:

Dr. Corrinne Caldwell, Advisory Chair, Educational Leadership
Dr. Joseph DuCette, Educational Psychology
Dr. Will Jordan, Educational Leadership
Dr. Jodi Levine Laufgraben, External Member, Temple University
Dr. Catherine Schifter, Psychological, Organizational & Leadership Studies in Education

ABSTRACT

This study investigated the relationship between the proportion of coursework students complete in an online format absent traditional classroom interaction and their expected score on a standardized, content-driven achievement test, holding constant other factors expected to influence test scores, including demographic variables, major, and prior academic performance.

The study's sample involved 817 undergraduate business majors at a small, specialized, private, not-for-profit U.S. institution of higher education and their performance on Educational Testing Service's Major Field Test in Business (MFT-B). Students in the sample chose course-by-course to take either a hybrid format that involved regular classroom meetings supplemented by online learning or an entirely asynchronous online course involving no face-to-face interaction. Learning outcomes, syllabi, assessments, duration and the pool of instructors were the same in both formats.

This investigation was motivated by the changing role of distance education in higher education and the increasing fungibility of credits earned at a distance and those earned in a traditional, classroom-based context. The use of the MFT-B was motivated by the growing emphasis on student learning outcomes assessment and mounting demands for "accountability" in higher education.

An ordinary least squares regression modeling MFT-B score as a function of proportion of credits completed at a distance, GPA, major, transfer credits, completion time, age, gender and ethnicity found proportion of study at a distance to have a strongly significant ($p<.001$), positive impact on expected MFT-B score. A logistic regression of

likelihood of graduation as a function of those same variables found a strongly significant ($p<.001$) negative impact of study at a distance on retention to graduation.

ACKNOWLEDGEMENTS

I am deeply indebted to Dr. Corrine Caldwell for all of her guidance and support, for the wisdom of her counsel, the inspiration of her example, and for all that I learned from her. I received great insight and direction from Dr. Joseph DuCette, and thank him for giving so freely of his time. I am indebted too to Dr. Patricia Rucker, a great boss and a great provost, for her unwavering support. I might not have made it through were it not for her words of encouragement, insight, and her patience when my research drew my time and attention away from the demands of my work. Thanks above all to my parents, John and Marsha, for their unconditional love, their belief that I could do anything I set out to, and the great sacrifices they made in the name of my growth, education, and wellbeing. Finally, special gratitude and love to my wife, Leah. She put up with a lot, to say the least, and I am so grateful to her for it.

DEDICATION

To Max and David,
that it might help inspire you to achieve your dreams

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1
INTRODUCTION

Technological innovation in education has been called a "cottage industry" (Bhattacharya 2008). Higher education is notoriously resistant to change of most sorts (Evans & Henrichsen, 2008; Lee, Hyman, & Luginbuhl, 2007). Yet online learning has experienced remarkable growth in the last decade; the current 17% rate at which online enrollment in higher education is growing far outstrips the just over 1% of higher education overall (Allen & Seaman, 2010). No consensus, however, has developed about how online learning's effectiveness, in terms of student learning, compares to traditional classroom methods.

The Spellings Commission Report (2006, p. 25) implores institutions to "use distance learning to meet the educational needs of rural students and adult learners, and to enhance workforce development. Effective use of information technology can improve student learning, reduce instructional costs, and meet critical workforce needs." Online learning's power to enhance access to higher education is surely beneficial. Allen, Seaman, and Garrett (2007) provide evidence, however, that students studying primarily in the traditional classroom context or otherwise able to attend class in person are increasingly opting, in whole or in part, to study at a distance. In these cases in particular, where studying at a distance is discretionary, it is important to understand the extent to which the absence of classroom interaction in asynchronous distance education impacts student learning outcomes.

Statement of the Problem

As the proportion of higher education in the United States that occurs at a distance continues to grow, the importance of knowing how the expected student learning outcomes from distance education compare to those occurring in the traditional classroom grows as well. While there are a host of both optimistic and skeptical theoretical bases for online learning's ability to effect achievement, the question is ultimately an empirical one, particularly given online learning's widespread adoption. Indeed, over the past decade literally thousands of individual and meta-analyses of the effectiveness of distance education versus classroom instruction studies have been conducted. However, like the saw about economists, stack those studies end-on-end and one still can't reach a conclusion. For every one study favoring classroom instruction there appears to be another strongly favoring distance learning and two more showing no significant differences. Meta-analyses synthesizing results across multiple studies consistently show strongly heterogeneous results (Allen et al., 2004; Bernard, Abrami, Lou, Borokhovsk, et al., 2004; Y. Lou, R. M. Bernard, & P. C. Abrami, 2006; Sitzmann, Kraiger, Stewart, & Wisher, 2006), implying the effects reported in the underlying studies do not cohere. Criticism of the methodological quality of distance education research abounds (e.g. Davies, Howell, & Petrie, 2010). In sum, the question of distance education's relative effectiveness appears to remain an open one.

At the same time, never has the assessment of student learning been more central to higher education and its regulation (Shavelson, 2010). An increased focus among institutions and policymakers on what Banta (2007) calls "assessment for accountability" is providing an increase in the availability of objective, arms-length data on student

achievement. Such data may well provide the foundation for improved understanding of the relative effectiveness of distance education (DE) and classroom instruction (CI).

Significance of the Problem

During the 2006-2007 academic year, 66% of all Title IV funding eligible, degree-granting post-secondary institutions offered some form of distance education, for an estimated total of 9.4 million online course enrollments (U.S. Department of Education National Center for Education Statistics, 2006). During the same year, the NCSE reports that 32% of those institutions offered programs designed to be completed entirely online, an estimated total of 11,200 programs, over 7,300 of which were degree-granting. Allen and Seaman (2010) report that in fall of 2008 over 4.6 million students were enrolled in at least one online college course, a 17% increase over the previous year, which well exceeds higher education's overall annual growth of 1.2%. Much evidence supports the belief that DE's impact and prevalence will continue to grow in higher education in the coming years. Indeed, 74% of chief academic officers at public institutions surveyed by Allen and Seaman reported believing that online education is critical for their respective long-term strategies.

At the same time, however, skepticism regarding DE abounds. Some of this skepticism may have its roots in the historical conflation of online education and for-profit education, and some in its historical association with correspondence courses. However, as public and private not-for-profit institutions increasingly offer DE, the two are becoming increasingly distinct. Within their respective institutions, faculty skepticism appears to be strong: in their survey of chief academic officers, Allen and Seaman (2010) found that less than one-third reported that their faculty accept the "value

and legitimacy" of online education. The proportion of faculty reported to accept online education varied significantly but in no case exceeded 50%. Further, these acceptance results have been fairly stable since Allen and Seaman began doing the survey in 2002.

Those responsible for hiring college graduates appear to share similar reservations. In a survey of 269 hiring executives involving the employability of three hypothetical candidates – one having studied entirely online, one partially, and one not at all, Adams and DeFleur (2006) found those who had earned a bachelor's degree either partly or entirely online were markedly less likely to be hired. In a qualitative study of hiring managers representing a broad set of industries, Seibold (2007) also found evidence that, despite online education's increasing visibility and growth, the perception persists among hirers that traditional degrees are superior to those earned online.

Purpose of the Study

The purpose of this study was to examine the relationship between scores on Education Testing Service's Major Field Test in Business (MFT-B), a standardized, objective exit examination, achieved by undergraduate business students at a small, specialized urban college and their percentage of coursework completed online, holding constant relevant demographic factors and other likely predictive factors such as GPA. An extensive empirical literature addressing the issue of DE vs. CI student performance exists, but one with widely uneven methodological rigor, a wide and perplexing field of theoretical bases that stand to inform expectations regarding the question, and a clear absence of consensus. This study's position relative to the literature is detailed in the following two sections.

"Hybrid" classes and the changing roles of technology and classroom interaction.

That the literature overwhelmingly compares technology-rich distance education to classroom instruction largely or entirely free from technological enhancement is unsurprising given that technology's impact in the classroom has long been of interest. Indeed, in many DE/CI comparative studies, courses involving significant technological mediation together with classroom instruction are classified as distance education for the purposes of comparing DE to CI. As the technology of distance education is an increasingly commonplace supplement to traditional classroom-based courses, however, studies investigating the impact of technology rich versus technology starved educational experiences become moot. The relevant question is fast becoming one of the relative effectiveness of a course based in the classroom that has all of the technological features of a modern learning management system versus an entirely asynchronous class with *only* those features and no classroom interactions. The impact of the presence or absence of a technology-rich environment is being supplanted in practical consequence by the impact of the presence or absence of substantive, synchronous, collocated interaction.

While numerous studies examining "blended" or "hybrid" education in the United States involving both the technological tools of DE and some classroom instruction exist, they typically lump blended learning in with distance education devoid of classroom interaction (e.g. U.S. Department of Education Office of Planning Evaluation and Policy Development, 2010), focusing not on the presence or absence of same-place, same-time interaction, but rather on the presence or absence of technology. (Gebara (2010) is a recent and noteworthy exception, although she uses GPA as a measure of academic achievement, which seems problematic on its face.) Again, though, this is not the trade-

off faced by many students in higher education today, who take traditional CI supplemented with DE-style technology or alternately take largely or entirely asynchronous, distance-only courses.

Allen, Seaman, and Garrett (2007) report that in 2004, 79% of U.S. public institutions offered one or more blended courses, and this figure excludes what they call "web facilitated" face-to-face courses that use a learning management system to "deliver" content, but for which the proportion of content so delivered is less than 30%.[1] In sum, as distance education has become more widespread, its tools have penetrated the traditional classroom as well. Many face-to-face, classroom-based courses are in fact blended, hybrid, or web facilitated classes. In many institutions, this fact is sufficiently standard operating procedure that such classes are not explicitly identified as hybrid or blended but rather simply as traditional, face-to-face classes. There is every reason to expect that this trend will continue, as learning management systems further pervade higher education.

In terms of market demand, the desire for blended programs also appears to be growing. In a survey of 2,033 prospective college students, Eduventures Learning Collaborative for Higher Education (2006) found that 22% of respondents preferred a totally on-campus program, 19% preferred a totally online program, and that the remainder preferred some form of blended program, with 25% wanting a mainly on-campus blend, 18% an online-centric blend, and 14% wanting an equal balance.

In light of these trends, the impact of the *technology* of distance education is not the variable of interest, because its presence is commonplace and growing. Rather, the

[1] The ostensibly difficult question of how the exact percentage of content delivered online versus in the classroom might itself be determined is left unaddressed.

presence or absence of *classroom interaction* is, particularly as students increasingly blend DE with CI. The study presented here, unlike the bulk of the DE-CI comparative literature, investigated the impact of classroom interaction while holding constant the availability and use of a set of learning management system tools, thus more closely reflecting the real tradeoff faced by students choosing between distance and traditional classes.

The "Blended" Student.

Traditionally DE-CI comparative literature involves a between-groups design. Indeed, a common methodological criticism of these studies is, among other things, the absence of random assignment of participants to the "treatment" and "control" groups (e.g. Bernard, Abrami, Lou, & Borokhovski, 2004; Ungerleider & Burns, 2003). While this is an entirely orthodox and valid methodological criticism, it puts researchers of all-or-none DE in a bind. One can study those naturally enrolled in a distance education program compared to their counterparts predisposed to the traditional classroom, but in so doing suffer potential selection bias through the natural differences between the groups. Alternately, one could randomly assign participants to DE or CI groups, although the relevance of the insight so gained is constrained by the fact that how well those who would not otherwise seek to learn online actually succeed in learning online is of academic interest at best.

The good news here is that the challenges of between groups assignment are in many instances becoming moot. Allen et al. (2007) report that in a 2006 Eduventures study 51.5% of those in postsecondary education have experienced at least one online or blended course. Thirty-one percent had experienced a blended course, and 23% reported

having been in a blended program, where students take some courses online and others in the classroom. Students who have never experienced an online or hybrid course are the minority; in the future one can expect them to be rarer still.

The study herein addresses these issues in the following ways: it involves students in a hybrid program, who opt on a course-by-course basis to take classes online or on campus. Some students take courses wholly online, others wholly on campus, but most opt for a blend of on-campus and online courses. As such, group assignment concerns are obviated per se and the potential for confounding variables is reduced. Because students are encouraged to pick and choose between online and classroom courses, the courses are by design scrupulously interchangeable, with the same outcomes, same texts, same assignments, and the same pool of instructors. Finally, the classroom-based sections are systematically supplemented with all of the same learning management system tools the online courses have, and these tools are used in the same fashion in strictly online and classroom-based sections (although naturally not to the same extent). These circumstances create a natural experiment of sorts, through which the impact of classroom interaction can be estimated more directly than in previous studies.

Research questions

The questions addressed in this study include the following:

- Research Question 1: Holding all other known relevant variables such as GPA, age, race, gender, major, and semester standing constant, does the extent to which students in a bachelor's program in business study at a distance as opposed to in a hybrid

format involving traditional classroom interaction impact their performance on the Major Field Test in Business (MFT-B)?

  - Sub-question 1: Do demographic characteristics such as age, gender or race change the way and extent to which proportion of online study impacts performance on the MFT-B?
  - Sub-question 2: Are there individual courses whose grades are, all else equal, more closely correlated with MFT-B scores than others?
  - Sub-question 3: In multi-course sequences (statistics 1 and statistics II, for example), does a student's having taken the first course in the sequence in DE or CI format impact the grade in the follow-on?

- Research Question 2: Holding all other known relevant variables such as GPA, age, race, gender, major, and semester standing constant, does the extent to which students in a bachelor's program in business study online as opposed to a traditional classroom impact their retention to graduation?

Definitions

Distance Education and Related Terminology.

Common terminology and operational definitions continue to elude the realm of distance learning / distance education / online learning / e-learning (Lowenthal & Wilson, 2010; Moore, Dickson-Deane, & Galyen, 2011). Distance education is the most common, least controversial and broadest nomenclature. Keegan's (1996) definition of distance education (DE), identified by Bernard (2004) as the most widely cited, identifies five key elements of the definition that distinguish DE from other forms of instruction: that teacher and student are separated at least semi-permanently; that an educational

institution is involved in the provision of student support services, technological mediation, provision for two-way communication; and the "quasi-permanent" absence of learning groups (Bernard, p. 380). The fifth criterion in Keegan's definition -- that of precluding learning groups -- has been, as Bernard notes, controversial, and as technological mechanisms for supporting learning groups at a distance both synchronously and asynchronously, is increasingly problematic. Part of the inconsistency in terminology and usage is as a consequence of a rapidly changing state of practice over time, particularly with regard to technological mediation. Even at a given time period, distance education practice is not one-dimensional, and practices, standards and tools vary widely (Bernard, Abrami, Lou, Borokhovski, et al., 2004).

DE is not a medium of education, but rather a means of education that depends upon the availability of media in a way that classroom education need not.

Synchronous versus asynchronous DE.

An important part of defining DE is to delineate the mechanisms and modalities by which interactions happen. By definition such interactions happen at a distance and mediated by technology, but within such circumstances they may occur in markedly different ways. Chief among these differences is whether interactions occur primarily synchronously, in real time (through, for example, a simulcast of an "originating site" – typically a classroom – being broadcast over the web or to one or more remote sites), or whether they occur primarily asynchronously, with students working independently and interacting with instructors via discussion board postings and email.

Much of the early funding of DE initiatives was done by the Sloan Foundation's Asynchronous Learning Network (ALN). Doherty (1998) provides a broad overview of

the genesis of Sloan's ALN approach. As the presence of the term 'asynchronous' in the program's very name would indicate, that a student could study at any time and from any location was pivotal to the Sloan approach to DE. Sloan was the primary funder of early DE initiatives (Blumenstyk, 1998) and perhaps as a result, most early DE projects were scrupulously asynchronous. The ALN approach had solid theoretical underpinnings: Doherty points to learner control theory (LCT) as a theoretical grounding of the ALN approach. Learner control theory posits that learners who are afforded control of the nature and context of their own learning will experience better outcomes than those for which those features are more closely prescribed by teachers or others. At least as early as Newman (1957) the positive impacts of learner control have been empirically demonstrated. How learning activities are undertaken, as well as when and for how long are key decisions best left, according to LCT advocates, for students to determine for themselves. Asynchronous learning is very much in keeping with this notion. Of course, the ability to broaden a course's reach across many time zones is also a function of the extent to which it remains asynchronous as well, and practical consideration were and continue to be a factor shaping the nature of DE. The Council for Higher Education Accreditation (CHEA) (2000) reports that most early DE was asynchronous, citing the available alternatives as one-way or two-way video-based technologies. A wide and flexible array of Internet-based synchronous technologies now avails itself (see Lavooy & Newlin, 2008, for an up-to-date overview of these options, as well as some empirical indication of their relative effectiveness). Modern practice allows a blending of synchronous and asynchronous interactions, for example through chat sessions,

teleconferences and live video simulcasts enriching an otherwise primarily asynchronous course.

An emerging standard of DE practice.

Techniques and practices vary in DE just as they vary in CI; the media by which teaching and learning transpires, while not entirely orthogonal to those standards and practices, is not the same as those techniques and practices. These facts notwithstanding, however, over time a rough standard of DE practice is emerging in higher education. The standard is for primarily asynchronous, web-mediated instruction that supports N-way interaction among instructors and students and that occurs typically via a learning management system. Learning management systems themselves are becoming increasingly standardized as the market is increasingly dominated by a handful of vendors. Over time, a standard of conventional practice in DE can be seen emerging.

Taylor (2007) develops an account of the historical "generations" of distance education practice. The first generation is the correspondence model; the second is asynchronous, largely non-interactive multimedia such as lessons captured on audio-or videotape. The third is marked by synchronous interaction of the sort often characterized as "tele-learning", that allowed students not physically present in the classroom to view a classroom lecture at a distance. The first three generations are now generally seen as outmoded forms of DE practice. The fourth generation is marked by varied modes of interactivity. It is primarily asynchronous, offers two-way communication between students and teachers, is Internet-mediated and may offer interactive multimedia content. The fifth and latest generation involves standardizing the learning management system and integrating it into the larger sphere of the institution's electronic resources and

beginning to rely on computer-mediated communication to provide automated interaction. While automated interaction remains marginal (and is likely not seen as a uniformly attractive proposition), the flexible, learner-directed, Internet-mediated DE of Taylor's fourth-to-fifth generation, with its focus on a variety of methods for student-student and student-teacher interaction and for student engagement with the course content, has the hallmarks of an emerging standard, and it is this DE that is the focus of this investigation.

Online learning.

Online learning is typically defined as educational circumstances whereby the learner accesses learning experiences mediated by some technology (Carliner, 2004). Increasingly the technology in question is the World Wide Web, and thus online learning is increasingly synonymous with web-based learning. While a precise definition of "e-learning" can be difficult to pin down (Moore et al., 2011), e-learning, online learning and web-based learning will be treated as synonymous here, each reflecting a modern, interactive, web-mediated, largely asynchronous form of distance education.

The Major Field Test in Business.

The Major Field Tests (MFTs) are "comprehensive undergraduate and MBA outcomes assessments designed to measure the critical knowledge and understanding obtained by students in a major field of study." (Educational Testing Service, 2011b). MFTs are discipline-specific multiple choice tests offered in several fields including business, the domain of interest here. MFTs purport to assess students' "factual knowledge, … ability to analyze and solve problems, understand relationships and interpret material from their major field of study" (Educational Testing Service, 2011b).

ETS administers the MFT-B at the associate, bachelors, and MBA levels. ETS lists 685 institutions that administer the bachelor's degree MFT in business to at least five students each. ETS provides comparative data on scaled scores of all seniors taking the test at U.S. institutions.

MFT reliability and validity.

Repeated application of a reliable instrument will yield consistent results. Educational Testing Service reports a measure of internal consistency validity for the MFT known as the Kuder-Richardson Formula 20. This formula uses the number of test items, the percentage correct and incorrect for each item and the variance in scores. The range of scores is from 0 to 1, with a high coefficient indicating an instrument more likely to correlate with alternative assessments. The Kuder-Richardson score for the MTF in Business (MFT-B) tests administered to college seniors from August 2006 to June 2009 was .89 (Educational Testing Service, 2010). Ebel and Frisbie (1991) indicate that Kuder-Richardson scores above .65 are adequate for inferences about groups and above .85 are sufficient for individuals.

ETS uses national curriculum surveys to specify the content of the MFT tests and employs content experts from both MFT-using and non-MFT-using institutions to provide content validity. Criterion or predictive validity is the extent to which a given measure is correlated with some other external measure. One would expect MFT-B scores to be correlated, for example, with major GPA. In an investigation of MFT-B criterion validity Allen and Bycio (1997) found the criterion with the strongest association was business course GPA. Mirchandani, Lynch and Hamilton (2001) found

that a GPA from a subset of quantitative courses (including calculus, accounting, finance, operations management and MIS, but interestingly, not statistics) best correlated with MFT-B scores.

Delimitation and Limitations of This Study

This study measures the correlation between MFT-B scores for students majoring in business administration at a small, private, specialized urban college and the percentage of their credits taken in asynchronous DE form, accounting statistically for other variables likely to be correlated with MFT-B performance, including major GPA, the student's choice of concentration, and demographic variables including gender, age, and race. Variables related to student or instructor satisfaction, student persistence, or the factors related to student preference for DE versus CI are excluded here to focus exclusively on an objective content knowledge outcome obtained by students who persist through to their senior capstone class, in which they sit for the MFT-B test.

Given that this study was limited to a single institution, it may not lend readily to generalization beyond that organization's walls. Given a potential disparity between DE and CI in retention, DE may benefit from a disproportionate "survival of the fittest" boost that may be difficult to control for. While I developed a model to estimate distance education's impact on retention, the role of retention disparity in influencing outcomes achievement differentials remains uncertain. The absence of a solid variable to capture general test-taking aptitude is a limitation warranting discussion here. General Education GPA is the best available candidate. It, along with major GPA and MFT score, are likely all underpinned by a more general underlying intellectual capacity. Such a capacity is likely to be independent of subject matter mastery or of the extent to which teaching and

learning is impactful generally. A proxy for such capacity in the model presented herein is desirable for this reason.

Significance of This Study

Many DE-versus-CI studies compare hybrid/blended learning where learning activities occur partly online and partly in the classroom against courses occurring exclusively in the classroom. That these comparisons tend to favor the hybrid courses is not particularly surprising. Nor does this kind of study inform especially well the most important DE versus CI policy questions. In particular, the issue of what if anything is lost or gained by way of learning outcomes when students opt to study at a distance instead of face-to-face is lost. Instead of examining the impact on outcomes of the presence or absence of learning management system technology – an increasingly irrelevant question as such technology continues to pervade traditional classroom-based learning – this study examines the impact of the same-time, same-place learning experiences that happen in the physical classroom. Is there evidence that such interactions enhance learning outcomes as measured by a standardized test of content knowledge?

Another source of significance is that a substantial subset of the participants in this study (around 50%) take a significant portion (no fewer than 20/80 DE-CI and no more than 80/20) of their coursework in both DE and CI formats, to some extent obviating the issue of random group assignment and helping address the issue of controlling for ability, etc. Instead of seeing DE versus CI as strictly an either/or, one can measure the impact of the *extent* to which coursework is completed online has on MFT-B achievement. This also allows for pre-post grade changes from one course in a multi-

course sequence to the other (Accounting I and Accounting II, for example) to be compared first course DE, second CI versus first course CI, second DE.

Theoretical Base

Activity theory (AT), the main theoretical lens for this study, is a psychological meta-theory developed principally by Leont'ev (1978) and Vygotsky (1994). Activity theory examines human cognition and action as understandable only as situated in the socio-technical context in which they occur. It recognizes that:

> Cognition does not exist outside the life process that in its very nature is a material, practical process. The reflection of reality arises and develops in the process of the development of real ties of cognitive people with the human world surrounding them; it is defined by these ties and, in its turn, has an effect on their development. (Leont'ev, 1978, p. ch1.htm)

In Leont'evian activity theory, activities are undertaken collectively by *subjects* – individuals and groups of individuals – motivated by the desire to transform *objects* (either physical or ideal) mediated by *tools* (either physical or symbolic representations, and, often in context at hand, complex information systems comprised of both). Engeström (1987) supplements Leont'ev's trinity of subject, object, and artifact with roles for the larger community, the division of labor within that community and the rules that structure both the community and the division of labor therein.

Activity theory is one among a set of theories that study cognition, action, social relationships and communication channels, together with physical environment and artifacts, as holistic systems. Other such theories include distributed cognition (Hutchins, 1995) and the situation/action perspective exemplified by Suchman (1987). Activity theory is the widest ranging and has attracted the most attention in the education

literature. Indeed, Blin and Munro (2008) employ activity theory as a lens through which to understand the lack of transformative impact e-learning has had in higher education. They argue after Engeström (1987) that DE is disruptive and not yet "expansively transformative" in that disruption. It has disrupted traditional teaching and learning structures, but it has not yet led to the synthesis of entirely new curricula, assessment techniques, pedagogies, patterns of work, or resources. As such, it has only disturbed the status quo and has been met mostly with resistance.[2]

Issroff and Scanlon (2002) use AT as an analytical framework through which to view several case studies involving distance education. Kerres and Witt (2003) take up the issue of appropriate "didactical design" from the AT perspective, while Oliver and Trigwell (2005) ground a discussion of the utility of the "blended learning" terminology itself in AT.

Activity theory also brings a valuable sophistication to the way in which technology's role in and consequence for education are assessed. Through AT's lens, seeing the medium by which education is "delivered" as inherently inconsequential and altogether beside the point (Clark, 1983, 1994, 2000) or alternately as thoroughly transformative (Smith & Dillon, 1999) – a common debate among those inclined to argue either in favor of or against the DE-CI comparative studies such as this – is when framed by AT grossly reductive. Debate on the respective roles of medium, method, and message ignores the systemic whole so important to AT. Learning results from interactions with others, mediated by language and other artifacts, technological and otherwise. AT conceptualizes learning "not as the internalization of discrete information

[2] Aggregate adoption rates –at the time of their writing, and well prior – surely might suggest otherwise.

or skills by individuals, but rather as expanding involvement over time, social as well as intellectual, with some other people and the tools available in their culture" (Russel, 2002, p. 65). This expanding social and intellectual involvement over time is not only the end result of having learned, it is the means by which learning transpires as well. Rather than debate the consequence of the media through which the knowledge generated by teaching and learning strategies is transmitted, one needs to see the class (in the broadest sense of the term, whether virtual or physical) as a complex sociotechnical system. Learning is social action mediated by artifacts (whether they be physical tools, digital artifacts of networked learning or the symbolic artifacts of language) in context. In AT terms, to understand learning or to assess the capacity for learning to happen requires understanding the "activity system" in which learning emerges.[3]

In Engeström-ian (1987) AT, an activity system involves a *subject* (learner) pursuing some *object*, or motive (learning about the subject at issue in a given course for example), as mediated by *tools* (discussion boards, books, PowerPoint slides, videos) to achieve particular *outcomes* as part of a *community* in which there is a *division of labor* and *rules* (norms, values). This activity system is in AT the basic unit of analysis. While one might increase the magnification of one's AT lens to focus more on one part of this system at the partial exclusion of others, the integrated whole still needs to be the object of inquiry. Isolating elements like the transmission of knowledge from teacher to student

[3] Admittedly, given that learning viewed through AT is the ability to engage in meaningful activity, supported by the appropriate set of tools, to achieve membership in a given community, an objective test of content mastery like the MFT-B is a poor measure of learning to say the least. Nevertheless, seeing the test as a rough-and-ready measure of the skills and knowledge necessary to participate legitimately in the larger AT system that is the community of professional business practice seems entirely in keeping with the AT weltanschauung.

across a medium from the rest of the activity system and its divers contexts is, seen from the lens of AT, misguided.

CHAPTER 2
LITERATURE REVIEW

Introduction and Organization

The literature review that follows consists of a brief introductory section exploring the literature that addresses questions regarding the validity and utility of the research questions proposed here, followed by a review of two primary relevant empirical literature strands. The first is a review of the extant empirical distance education versus classroom instruction comparative literature. Given the nature of the research here, the need for such a review is self-evident. Given the enormous volume of this literature, the review is constrained to meta-analyses – quantitative research whose unit of analysis is other published studies and whose results synthesize effect sizes across studies. A review of these summative meta-studies provides a coherent and manageable view of a corpus that without which would be impracticable to characterize.

The second empirical strand involves the research on modeling student standardized test performance and particularly performance on the MFT-B. Given that the model employed here uses the MFT-B as its dependent variable, a thoroughgoing explication of the literature on factors influencing such scores is critical as well. Assessing the extent to which these strands have intersected is important for understanding the novelty of the research here. To that end, this review plumbs the DE-versus-CI literature for instances where the MFT has been used as an indicator of student achievement and also the psychometric literature, for predictive/explanatory models of MFT performance that have examined the impact of distance education.

First Main Thread: Comparative Distance Education Studies

While DE has become sufficiently pervasive that students of all stripes and backgrounds have studied at a distance, the demographics of DE students do differ from those of traditional, on-campus students. The GAO reports (*Growth in Distance Education Programs,* 2002) that online students compared to their traditional counterparts tend to be older, to work full time, to study part-time, and to be married. Policymakers seek better student access and responsiveness to changing demographics and thus favor DE: "We want postsecondary institutions to adapt to a world altered by technology, changing demographics and globalization, in which the higher-education landscape includes new providers and new paradigms, from for-profit universities to distance learning." (*A Test of Leadership*, 2006, p. xi) At the same time, DE clearly has its skeptics (Smith & Mitry, 2008, for example). The potential for DE to be in practice little more than a more technologically sophisticated, more widely accepted form of correspondence course is real.

Appana (2008) elicits the forces driving DE's widespread and rapid adoption. Chief among them is access to new markets, both in the form of broader geographical reach and in terms of appeal to student demographic groups such as working adult students either disinclined or unable to travel to campus. Next is cost savings through reduced need for physical classroom space and reduced variable costs generally. While cost savings is among Appana's list of benefits to institutions offering online classes, Smith and Mitry (2008) observe that it is not yet clear whether DE can be relied upon for reduced costs or financial success. They note that online programs at NYU, Wharton, Columbia and Temple all closed shortly after opening and suggest that online programs

are profitable only to the extent they engage in quality-compromising corner cutting. While the idea of a centralized, for-profit DE unit has fallen out of favor because of misestimated cost savings and enrollment potential, DE continues to thrive and grow at many institutions in a more decentralized, organic manner (Finkelstein, 2000; Smith & Mitry, 2008).

Does Comparing DE to CI Make Sense?

Certain DE researchers appear to find further inquiry into the effectiveness of DE relative to CI needless because of DE's widespread acceptance or otherwise a priori irrelevant (Brown & Wack, 1999; Saba, 1998). This question begging, while certainly pragmatic for scholars of distance education, should leave policymakers and educators unmoved. While technological adoption may generally be a one-way street, to the extent that students who could readily study in a classroom setting are opting instead to study at a distance, DE's broad acceptance does not make the issue of relative effectiveness moot.

A subtler but no less problematic basis for dismissing DE-to-CI comparisons is that they are somehow unjust and unproductive. Bernard et al. (2009, p. 1167), for example, conclude "that little more can be gained through comparisons between DE and CI". Their reason appears to center on the fact that "in DE versus CI studies, delivery method is often confounded with instructional design, in which the DE condition has instructional design features not present in the classroom control condition and vice versa". They are not wrong about this; in modern practice, media or "delivery method" surely is entangled with pedagogy and instructional design. Why would not or ought not it be, and why would this circumstance move one to call DE-to-CI comparisons unproductive or misguided? Two main mechanisms for education in the 21st century are

emerging. Their media and their pedagogy may be very different, and those differences in pedagogy may be in part an ineluctable consequence of their media. Either way, students are increasingly treating them as interchangeable. Nagel (2009) reports, for example, that the vast majority of college students will by 2014 take some of their classes online.

But do they achieve equivalent ends, and do they achieve them equally well? Educational equivalency is an enormously complicated and multifaceted construct, difficult to operationalize. Even more narrowly, operationalizing and measuring learning outcome achievement is notoriously difficult. Those challenges notwithstanding, as more and more postsecondary education occurs at a distance, the need to know is more imperative than ever.

The influence of access.

DE surely improves access to higher education, making it practical for those for whom college otherwise might not be. For many learners, particularly those who are rural or remote, as Bernard, Abrami, Lou, Borokhovski, et al. (2004, p. 416) note, "the appropriate comparison may be with 'no instruction' rather than 'traditional classroom instruction." Indeed, from this vantage of DE's enhanced accessibility and the archetypical distance-education-or-no-education case, comparing DE to CI effectiveness is misguided. Further, barriers to classroom access are by no means limited to the geographical and logistical. Financially independent, adult students also have jobs, families and related obligations to manage, to extents and in ways that can preclude regular classroom attendance.

However, traditional undergraduates and graduates alike increasingly have (and indeed take) the option to take one or more of the classes online in an otherwise "traditional", face-to-face program. Those who do so are obviously not opting to take a subset of their coursework online because they do not have access to the classroom. They may do so to more effectively manage their schedules, or otherwise for their own convenience, or because DE courses better align with their own learning styles, or for a host of other reasons, but given that they are in the classroom daily anyway, theirs is not a decision in this case driven by access – geospatial, schedule-wise, socioeconomic or otherwise.

Further, Eduventures Learning Collaborative for Higher Education (2006) reports that for many students in higher education studying at a distance, that distance is in fact quite modest, with over a third of all online students living within a 50-mile radius of the institution at which they are studying. Further, among those considering an online program, Eduventures Learning Collaborative for Higher Education reports that two-thirds of prospective students care about geography and 41% would prefer at least a branch campus of the institution they study at to be within 10 miles of their location. They conclude that a majority of the market for online study is local or regional. Moreover, the main reasons prospective students cite for desiring online programs with geographic proximity are the possibility of face-to-face contact and the ability to engage in blended learning (together with the ability to qualify for in-state tuition). While this does not preclude the existence of barriers for many students and potential students to studying in a traditional classroom context, it does further the argument that a large

subset of students in higher education are studying partially or entirely in DE that do so as a matter of choice, rather than as a matter of practical necessity.

Media's Role

The most salient difference between DE and CI is the nature and extent of the instructional media upon which they rely. Clark (1983, 1994, 2000) argues that media do not and will never influence learning. If so, then, as Simonson, Schlosser, and Hanson (1999) argue explicitly in their "equivalency theory", DE-CI comparisons are ultimately moot. If media has no impact, then DE-CI comparative studies can at best identify the impact of confounding variables. Others, though, argue that media is paramount, either because of its entanglement with message or because of the ways in which media constrains the social interaction central to learning. These arguments leave room for DE-CI comparison to be valuable. Each of these arguments will be considered in turn in the following.

One strain of thinking on the role of instructional media typified by Clark (1983, 1994, 2000) argues that the DE-CI comparison is moot because DE marks only a change in the media channel through which education is transmitted, and that such channels cannot impact the educational "payload" so carried. Clark (1983, p. 445) argues that

> "Media are mere vehicles that deliver instruction but do not influence student achievement any more than the truck that delivers our groceries causes changes in our nutrition. Basically, the choice of vehicle might influence the cost or extent of distributing instruction, but only the content of the vehicle can influence achievement."

If the technology that mediates DE is simply an alternative channel by which a payload of content and pedagogy gets delivered, then are inquiries into the effectiveness

of one medium compared to the other moot? Clark (2000) argues as much, but there are numerous bases for bringing skepticism to such an assertion.

For example, McLuhan (1964) famously asserts quite to the contrary that the "medium is the message", and that media can create an environment replete with consequences and constraints by its very presence. The notion that education can be modeled as the delivery of content via one or more media is itself subject to considerable criticism, most prominently by constructivists and particularly by those whose focus is on the social nature of knowledge construction (e.g. Vygotsky, 1978; Wenger, 1998) because the nature and scope of social interactions is clearly changed when interactions occur exclusively at a distance, as they must in DE, and particularly when they are occurring primarily asynchronously. Brown and Duguid (2000) envision a limited role for distance education in colleges and universities specifically because of the constraints DE places on the nature of socialization. Note, though, that the extent to which computer mediation of the sort occurring in modern DE can support the conditions necessary for learning as conceptualized by constructivists is very much an open question. Indeed, the discipline of "computer supported collaborative learning" has arisen around leveraging information technology to social constructivist ends (see Koschmann, 1996; Stahl, 2006).

In sum, theoretical support for the proposition that DE may differ systematically in its student learning effectiveness from that of traditional classroom methods is in ample supply, Clark's (1983) skepticism notwithstanding. Ultimately, there seems little reason not to allow the question of whether delivery channel makes a difference or not to be an empirical one, and empirically the evidence appears to show that the question is

still very much open, in spite of the enormous amount of investigation it has received. Such investigation will be reviewed in the following section.

Empirical DE/CI "Significant Differences" Studies

There have been literally thousands of research studies assessing the extent to which DE is different from traditional, classroom-based education. Russell (1999) was a seminal early cataloguer of such studies, assembling a collection of 355 research reports that found no significant difference between traditional and distance education. The variables under examination in the studies Russell catalogued varied widely, ranging from grade point average to persistence to student attitude to instructional resource utilization. In general, though, the research Russell collected sought to address whether taking a course in a distance education format reduced a student's chance of success. The distance education contexts so studied varied from literal correspondence courses to the asynchronous web-mediated courses typical of today's DE practice.

The methodological rigor of these initial "no significant differences" (NSD) studies varied widely. On his "nosignificantdifferences.org" website, Russell (2010) acknowledges significant threats to the validity of the NSD studies, particularly in lack of control for differences between the DE to the traditional student samples analyzed. The site continues to catalogue traditional versus DE student outcome studies that find both no significant differences and significant differences (both in favor of DE and in favor of traditional education).

The diversity and sheer volume of DE versus classroom instruction (CI) studies has given rise to at least sixteen studies (Allen, Bourhis, Burrell, & Mabry, 2002; Allen et al., 2004; Bernard, Abrami, Lou, & Borokhovski, 2004; Bernard et al., 2009; Bernard,

Abrami, Lou, Borokhovski, et al., 2004; Cavanaugh, 2001; Cavanaugh, Gillan, Kromrey, Hess, & Blomeyer, 2004; Cook et al., 2008; Jahng, Krug, & Zhang, 2007; Y. Lou, R. Bernard, & P. Abrami, 2006; Machtmes & Asher, 2000; Shachar & Neumann, 2003; Sitzmann et al., 2006; U.S. Department of Education Office of Planning Evaluation and Policy Development, 2010; Ungerleider & Burns, 2003; Williams, 2006) between 2000 and 2010 that endeavor to synthesize these findings through meta-analysis. Bethel and Bernard (2010) undertake a second-order synthesis of the synthetic DE-versus-CI literature, producing a typology of qualitative and quantitative synthetic approaches. Second-order meta-analyses, whereby meta-analyses are themselves meta-analyzed, appear in the literature on related questions (Barneveld & Strobel, 2009, on the application of problem-based learning; and Tamim, Bernard, Borokhovski, Abrami, & Schmid, 2011, on the impact of instructional technology in classroom instruction) but second-order DE-to-CI comparative meta-analyses do not appear to have entered the literature yet.

The DE-to-CI Meta-literature

The overwhelming multitude of primary research studies precludes practicable review herein. The prevalence of published meta-analyses and even second-order meta-analyses largely obviates the need for such an undertaking. Indeed, Heberlein and Baumgartner (1978, p. 448) praise the "clearer, more parsimonious review" that meta-analysis provides as compared to a qualitative analysis of the literature provided by an individual reviewer.

The Individual Meta-analyses

In the following section the second-order quantitative analyses of the primary empirical literature comparing the performance of DE compared to CI published between 2000 and 2010 will be reviewed in rough chronological order.

Machtmes and Asher's (2000) research, which largely predates web-based instruction, synthesizes studies comparing traditional CI to "telecourses" – courses whereby proceedings are captured for those not physically present, either by simulcast for students watching at a distance, or recorded live for subsequent review. Cavanaugh (2001) evaluates similar telecourses in the K-12 context. Allen (2002) examines student preferences as opposed to actual student outcomes. Shachar and Neumann (2003, p. 3) study college courses and use academic achievement as their dependent variable, albeit in the form of "course grades, tests, and exams." They acknowledge (p. 3) that "although course grades may carry some assessor subjectivity, they may be regarded as 'more objective' than other factors." While grades may well be 'more objective' than other dependent variables in the Shachar and Neumann's study typology – variables like satisfaction, attitude, or course evaluation – possible artifacts of the inherent subjectivity, particularly a potential tendency for grade distributions within sections to be explicitly or implicitly self-norming to some degree, raise significant questions regarding their validity as objective achievement measures. Further, while Shachar and Neumann indicate they maintained a study inclusion criterion related to methodological quality, they are not explicit about the nature of that criterion outside of excluding studies with "severe methodological flaws" (p. 5). Uncontrolled variables that differ systematically between DE and CI samples stand to confound outcomes significantly. If DE populations are

qualitatively different in ways that impact academic performance and that are not explicitly included in a study's model then basis for finding methodological flaw may exist.

Ungerleider and Burns (2003) improve on the abovementioned studies by limiting their study sample to what they call "networked education", excluding studies that involve only one-way, teacher-to-student communication and limiting their timeframe to 2000-2002. Their study was also scrupulously international, covering articles published in a wide array of languages. They found that only 44 of the original 135 studies of interest had a control group (that is, actually compared a group pursuing educational ends through technology- and distance-mediated means to those doing so in a traditional classroom). Of these, seven involved randomized assignment. In only two of the studies was rating group-assignment blind. Experimenter bias in investigations such as these is an entirely real threat; many studies come from unabashed DE boosters dazzled by technology's promise, others from those out to defend in the classroom a sacrosanct and timeless way of academic life. Ultimately Ungerleider and Burns found only 12 studies that met their inclusion criteria and had sufficient statistical detail for inclusion. They found no significant differences between DE and CI in terms of student achievement. Their conclusions are at least as much about the methodological inadequacies in the primary DE versus CI research as they are about their own empirical findings.

Allen, Mabry, Mattrey, Bourhis, Titsworth and Burrell (2004) examine studies whose variable of interest is student learning outcome effectiveness. They define (p. 406) effectiveness for their purposes as "performance related to scores on tests, grades achieved, or other similar evaluation of student performance" – measures that are readily

susceptible to the pressures of intra-sectional norming – but in the next sentence indicate that DE-to-CI comparisons were based on "either cross-sectional post-test-only comparisons or longitudinal changes score comparisons" (p. 408), implying that they might have limited their studies to those providing some form of course section- and delivery mechanism-spanning assessment measure that might be immune to intra-sectional norming.

In identifying candidate studies for inclusion in their meta-analysis, Allen et al. (2004) searched the relevant educational research indices via keyword but also included a manual search limited to all journals related to distance education. While the effort to include any published research meeting their inclusion criteria is otherwise laudable, the inclusion of journals specific to distance education raises the specter of publication bias. This bias may extend beyond the mere general tendency for studies showing statistically significant results to be published disproportionately often. The possibility exists that the journals themselves, given their self-interest in promoting the pedagogical mechanism that is their very reason for being, may be biased in favor of results supportive of distance education. Of course this issue is far broader than Allen et al.'s work. Much of the DE-CI comparative literature has been published in journals dedicated to DE.

Allen et al. (2004) found an overall average effect that shows a small improvement in student performance for distance education. The set of effects is strongly heterogeneous, however ($Q = 169.10$), indicating variation in effect sizes well beyond those expected by chance and the probability that one or more moderator variables are influencing the outcomes. In a search for moderator variables neither delineating between asynchronous and synchronous DE nor in channel of DE delivery provided

homogeneous results. Stratifying the studies by course content led in two out of six categorizations to homogenous results. Instruction in a foreign language showed a more strongly positive average effect size (r = .218) for distance education and homogenous results. Care in interpreting these results, as the authors themselves point out, needs to be taken, in that in all the studies examined, distance education was designed to facilitate interaction with native speakers of the language under study. This is a qualitatively different end to which to put DE than can be expected in other contexts. The effects reported in studies of education courses also showed homogeneity. The average effect size was negative but close to zero (average r = -.021), indicating a small average advantage resulting from traditional instruction in education over DE methods. Business education, the intended subject of this research, did not number among the course content areas affording further analysis. The potential for reduced heterogeneity seen in limiting studies to a given subject area may lend credence to subject-specific study such as the one herein.

Bernard, Abrami, Lou, and Borokhovski (2004) provided a comprehensive meta-analysis of the DE vs. CI literature from 1985 to 2002, and coded each candidate study on thirteen separate criteria, including publication type, type of measure (standardized test, researcher-made test, teacher-made test or researcher/teacher-made test), effect size, treatment duration, duration equivalency between experimental and control groups, instructor equivalency, student equivalence (via random assignment or statistical control), student time-on-task equivalence, instructional material equivalence, learner ability equivalence, dropout equivalence, class size equivalency and gender equivalence. Among all the occasions each of the studies analyzed had to report those criteria, 35% of

the time the information was unreported to readers, moving Bernard, Abrami, Lou and Borokhovski to lament not only the methodological rigor of the studies, but the methodological transparency as well. For each of the abovementioned study feature variables, for the subset of studies reporting that feature, the authors calculated the effect size and standard error for each value that variable can assume. For example, they found the average effect size for studies that do not use equivalent teaching materials to be positive (favoring DE), while those that do use the same materials to be negative (favoring CI), although in neither cases were the effects significantly non-zero. That pattern is repeated for same versus different time periods, instructor equivalence, and time equivalence, although the fact that effect size was in any case not significantly different from zero precludes any conclusions one might be tempted to draw. Note too that these feature effect size comparisons are done pair-wise, in isolation from one another.

In the final step of their meta-analysis, Bernard, Abrami, Lou, and Borokhovski (2004) modeled via multiple regression study effect size as a function of the set of aforementioned study features. In so doing, they sought to estimate the impact of study features exhibited in the set of studies making up their meta-analysis on the resultant effect size reported by those studies. For asynchronous DE, significant study features included gender equity, attrition equity, and longer courses, which all favored a pro-DE effect size, and class size equity, ability equity and time-on-task equity, which all favored a pro-CI effect size. These results warrant brief discussion in the context of the study conducted. Rather than assuring gender equity between control and experimental pools, investigating the effect of gender on DE vs. CI effectiveness – as this study has done –

may be more revelatory. That attrition equity appears to favor DE is interesting, but the notion of attrition equity is itself difficult, given evidence that retention is often more difficult in asynchronous DE courses (Bernard, Abrami, Lou, Borokhovski, et al., 2004; Xu & Jaggars, 2011a). Examining, as this study does, a circumstance where the same set of students switch freely between the CI and DE context, and who frequently take both, albeit to widely varying degrees, can obviate several of these concerns.

Bernard, Abrami, Lou, Borokhovski, et al. (2004) conducted a related meta-analysis with the same underlying study data. This meta-study included together measures of student achievement, attitudinal variables and student retention, although several results were reported per dependent variable category. While "distance from instructor" was a primary condition for inclusion among the experimental (DE) studies, they muddied the water somewhat by including among the DE hybrid courses composed of less than 50% face-to-face meetings, although they assure that "studies in which electronic media were used to supplement regular face-to-face classes with the teacher physically present" (p. 389) were excluded from the sample. Unsurprisingly given the time period under analysis (1985 – 2002), the DE sample included primarily or exclusively synchronous classroom simulcasts that are not representative of modern DE practice. Results, however, were reported separately for asynchronous, synchronous, and unclassified studies. For outcomes-related variables, the average effect size favored DE significantly ($k = 174$, $N = 36{,}531$, $g = .0527$, $SE = .0121$, $p<.05$), although with evidence of heterogeneity far beyond that expected by chance ($Q = 779.38$, $df = 173$). Average synchronous effect size favored traditional CI ($g+ -.1022$, SE .0236) with less extreme but still significant heterogeneity ($Q = 182.11$, $df = 91$). Regarding studies of retention-

related variables, the average effective size for asynchronous DE studies was small but significantly negative, with modest but still significant evidence of heterogeneous results. If asynchronous DE tends to lead to reduced retention rates, this alone could improve measures of student achievement. If only the academically prepared survive (and it seems reasonable to expect that they disproportionately would), then one could expect reduced retention to lead to high achievement among those who endure.

Again using the same underlying study data as Bernard, Abrami, Lou, and Borokhovski (2004) and Bernard, Abrami, Lou, Borokhovski, et al.(2004), but limiting the studies to the subset related to undergraduate classes, Lou et al. (2006) extend this prior work to focus on the interaction of media and pedagogy and to lend insight to the media-makes-no-difference (Clark, 1983, 1994, 2000) versus media-stands-to-transform (Kozma, 1994; Smith & Dillon, 1999) debate. Again asynchronous studies had a DE-favoring mean effect size that was small but significant, although again with distinct heterogeneity. Given the state of the practice in DE during the 1985-2002 time period and its diversity of technologies and technique, such heterogeneity may not be surprising:

> For example, some used Web-based resources, some used online course management systems with discussion board, some used Listserv, and some used broadcast TV or videotapes. The instruction in asynchronous DE was not simultaneously tied to a face-to-face host classroom, although in some cases real-time text-based chat may have been used for communication among students, or between students and the instructor. (Lou et al., 2006, p. 158)

Interestingly, regressing effect size on measures of methodological quality, opportunities for interaction, and demographic characteristics found significant, positive impact for the occasion for students to meet face-to-face both with one another and with the instructor, the very hallmarks of traditional classroom instruction.

In their meta-analysis, Sitzmann et al. (2006) searched both the training and the education literature for "web-based instruction" versus CI comparison studies during the period 1996 to early 2005 in an effort to understand the role of DE in workplace training. In addition to typical keyword database searches, they searched the *Journal of Asynchronous Learning Networks* manually and sought unpublished studies from the "Advanced Distributed Learning" electronic mailing list, identifying 96 studies to evaluate against further inclusion criteria, which included appropriate statistical reporting, student age of 18 or older, and the criteria that the training occurring was related to job knowledge or skill. (While they report that 67% of "trainees" were undergraduates and 18% were graduates, they do not operationalize training or comment on how or even if it differs from typical credit-bearing college courses.)

Sitzmann et al. (2006), apparently after the classification proposed by Kraiger, Ford and Salas (1993), attempted to separate those courses whose learning outcomes focused on declarative knowledge (i.e. the memorization of facts or principles) from those focusing on procedural knowledge (i.e. task performance and mastery). They coded studies for declarative versus procedural outcomes and found that for declarative knowledge, average effect size favored DE significantly while the average effect size for procedural knowledge favored CI but was not significantly different from zero, raising the prospect that different types of learning outcomes may yield different effectiveness differentials.

Sitzmann et al. (2006) did a sample size-weighted hierarchical regression of effect size against several moderating variables for the declarative knowledge subset. The mean age for both the DE and the CI groups were significant in the regression, with DE's

beta positive and CI's negative, meaning that as the average age of the DE students increased relative to that of the CI students, the increase in learning achievement expected for DE over CI increased as well. Age differentials alone explained 44% of effect size variance. Even broken down by moderator variable, effect sizes were significantly heterogeneous. For the subset of studies that involved random student group assignment, the positive effect of DE was reversed and CI was 10% more effective than DE. DE as a supplement to CI (as in "blended" or "hybrid" courses) was significantly more effective in teaching both declarative and procedural knowledge compared with CI alone (although such studies were sufficiently rare to preclude within-group analysis of moderator variables), a result the authors indicated warrants additional research. Degree of learner control (of pace or sequencing, for example) was correlated with a larger positive effect size for DE, suggesting that enabling learners to take control of the nature of their learning may be more consequential in the online environment.

Williams (2006) conducted a meta-analysis of 25 studies published from 1990 to 2003 comparing DE and CI in higher education in the allied health professions. Williams' study inclusion criteria were less restrictive than previous studies reviewed, and focused mainly on availability of the statistics needed to conduct a quantitative meta-analysis; nevertheless, only 19% of the 131 studies initially identified qualified for inclusion. The average effect size for student achievement favored DE to a small but significant degree. Student achievement, however, was assessed primarily by course grades. Course grades are a problematic measure of learning outcome achievement (Hu, 2005; Suskie, 2009), particularly in the absence of the relatively rare circumstance of the common examination (Bond, 2007). While its importance may be compromised by this

achievement variable validity issue, one noteworthy result noted by Williams is that working professionals experienced significantly better results with DE relative to CI than traditional aged students did, a result consistent with Sitzmann et al. (2006). Limiting the analysis to the subset of studies related to asynchronous DE resulted in a very small negative effect size.

Cook et al. (2008) also examined DE in healthcare, focusing on practicing and student physicians, dentists, nurses, and pharmacists, compared to traditional instruction. They found small, heterogeneous effects. Unlike any of the other studies in the literature, the authors also analyzed the literature comparing Internet-based learning to control groups involving no intervention at all, finding strong evidence that DE is in fact better than nothing, although the very existence of a literature investigating that question seems as much an indictment as anything, the faint praise of a positive result notwithstanding. Cook et al. found that outcomes measuring student skills and especially student behaviors in practice and their effects on patients tended to favor CE more so than knowledge-oriented outcomes. The wide range of study effect sizes reported in the literature analyzed and the marked heterogeneity in the average effect moved the authors to call for further research of the most effective means of DE and the most effective contexts for its use. They also conclude (p. 1195) that there is "limited value in further research seeking to demonstrate a global effect of Internet-based formats across learning, content domains, and outcomes", a recommendation that is amplified by the marked heterogeneity reported in virtually every meta-analytic comparison study reviewed.

Bernard et al. (2009) addressed Cook et al.'s (2008) call for a better understanding of the nature of effective DE through a meta-analytic investigation of the

importance of different categories of interactions involved in various DE "instructional treatments". They advocate an end to DE-to-CI comparative studies to enable better focus on what sort of DE works best. To that end, they conducted a meta-analyses of the literature comparing one type of DE intervention to one or more others, each with differences in their respective "interaction treatments". Interaction treatments associated with student engagement – either with one another, instructors, or course content – were found to be strongly predictive of student achievement, particularly for entirely asynchronous courses relative to those that involved face-to-face or other synchronous interaction.

The U.S. Department of Education Office of Planning Evaluation and Policy Development's (2010) DE-to-CI comparative meta-analysis stands to be conclusive. The quality of the underlying primary research may be seen to have precluded that, however. Starting with over 1,000 studies, only 50 effects reported in 45 studies met inclusion standards and reported the statistical information necessary to calculate an effect size. Most of these were published after 2004 and thus have reasonable currency. Of the 50 effects, 27 compared exclusively online experiences to CI, while the remaining 23 compared blended/hybrid classes to CI. While they did report the results of these two categories separately, the author's main finding that students in online learning "performed modestly better than those receiving face-to-face instruction" (p. ix) presented in the report's abstract and executive summary – depends entirely on the 23 blended-to-CI-comparisons to hold.

Unlike the study presented here, the comparison mounted by these studies is of technology rich instruction involving face-to-face interaction to technology-starved

instruction involving face-to-face interaction, an increasingly irrelevant comparison as the 21st century wears on.

Limiting the analysis (U.S. Department of Education Office of Planning Evaluation and Policy Development, 2010) to only those studies that compared online-only DE to CI resulted in no significant differences. No effort was made to limit studies to those controlling for time-on-task, instructional materials, nor did failure to control for retention rates preclude studies from inclusion. And while the report noted the potential for bias resulting from the author's dual role as subject/instructor and researcher, they did not preclude such studies from their meta-analysis. In neither of the sub-categories (blended versus CI or online-only versus CI) were the effect sizes homogeneous. The authors conducted an analysis of potential moderator variables explicit in a subset of the candidate studies, but did not break these results out across the hybrid-CI / online-only-CI categorization (likely because of insufficient data) to consider this categorization as a moderator variable itself. Few additional moderator variables appeared to differentiate effect sizes significantly.

One interesting result from the U.S. Department of Education Office of Planning Evaluation and Policy Development's (2010) report is that while individual studies that compared blended learning to exclusively online learning usually found comparable learning achievement across the two, the average effect size of studies comparing hybrid learning to CI favored the technology-enhanced classes more than online-only DE-to-CI comparisons did. If hybrid and online-only contexts tended not to differ in individual studies, one would expect that when compared to CI they would fare roughly equivalently well there too. Nevertheless, the increased impact of hybrid courses relative

to online-only courses, reported also in Sitzmann et al. (2006), may hint that face-to-face contact has a positive impact on student learning.

Koenig (2009), while he found student satisfaction significantly better in classroom delivery than online, also found positive themes from students regarding both classroom and online instruction in his qualitative analysis. The classroom was reported to afford students better concentration, allow greater student-instructor interaction, and enable physical actions and representations more effectively. Online delivery, meanwhile, was reported to encourage additional participation and allow self-paced learning.

Conclusions from and Summary of the DE-CI Comparative Literature

Overwhelmingly the primary and secondary quantitative studies of DE and CI in the literature make a comparison that is less and less the relevant question as technology continues to pervade higher education. They compare instructional contexts that involve technology – irrespective of whether or not they also involve classroom instruction – against those that are technology-free. However, traditional classroom instruction entirely free from any supplementation by at least some subset of the functionality of the increasingly ubiquitous learning management system, while by no means unheard of, is certainly likely to become increasingly rare.

This literature provides strong evidence that the technology tools of DE can be useful in improving learning outcomes, and, given their often considerable expense, that result is undoubtedly heartening. Studies comparing blended instruction in which the classroom is supplemented with DE technologies and interaction to technology-free classroom instruction (see Sitzmann et al., 2006; U.S. Department of Education Office of

Planning Evaluation and Policy Development, 2010) report, rather unsurprisingly, that better outcomes result with classes enhanced with technology.

Technological support for a traditional classroom course, often by the very learning management system through which DE courses themselves occur, is quite common currently. A not-too-distant future in which all classroom instruction is 'blended' is not difficult to image. Thus for policymakers and educators alike, the real question of interest is not the impact of increasingly ubiquitous technology but rather the consequence of the face-to-face contact that is becoming a less and less common component of higher education. Among the empirical comparisons should be hybrid face-to-face courses (again, increasingly the only sort of face-to-face courses that exist) against face-to-face-free, online-only courses. It is through this comparison that one may understand what if anything is lost in terms of learning achievement in relinquishing classroom interaction. Better understanding the consequences of the trend away from classroom interaction calls out for a multitude of methods, perspectives and approaches, many of them qualitative (e.g. Chyung & Vachon, 2005; Priebe, Ross, & Low, 2008). A quantitative investigation of the potential impact of classroom interaction on student learning outcomes would add to the literature as well.

Summary of the State of the Empirical DE-to-CI Comparative Literature

In a review of trends in North American DE scholarship, Davies, Howell and Petrie (2010) laud the decrease in DE-to-CI evaluation studies they find in the literature from 1998 to 2007: "On a positive note, projects directly comparing distance education with traditional face-to-face classrooms to determine the merit of specific programs declined significantly…" (p. 42). Several reasons might undergird such approval. If the

matter of relative merit were empirically settled in a conclusive way, then one certainly might cast the decline in such evaluations in an approving light. Davies, Howell and Petrie themselves, however, acknowledge critics of the methodological quality of such research, however, and the "questionable quality of assessment instruments used to gather comparison data" (p. 43) in a way that would likely preclude seeing the matter as largely or even preliminarily settled empirically.

Another factor that might move one to approve of the decline in comparative evaluation studies out of hand is an implicit desire to see the issue go away. Publishing their work as they do in the *International Review of Research in Open and Distance Learning*, Davies, Howell and Petrie (2010) (along, one might suspect, with their editors, reviewers and readers) may well be comfortable assuming implicitly that distance education is as good or better than traditional classroom alternatives. Indeed, they seem to make no bones: "…the need to validate the importance and viability of distance education based on comparisons with face-to-face learning experiences seems to expose a deep-rooted insecurity within the distance learning community – a fear that distance education is regarded as a somewhat substandard and less valued educational practice." (p. 44). The less parochial concern is surely not that a fear among distance education researchers and practitioners that their pursuit might be seen as second class but rather whether such regard (or lack thereof) has basis in evidence. That the literature on distance education takes the utility and role of distance education as given is neither surprising nor unreasonable. It provides, however, a truly shaky foundation for policymaking.

Specific Methodological Issues in the Empirical DE-CI Comparison Literature

The following section details methodological issues identified in the empirical DE-CI comparison literature and discusses their relevance to this study.

Bernard, Abrami, Lou, and Borokhovski (2004) reviewed the literature on DE-CI comparison literature and detailed elements that they understatedly characterize as "amenable to improvement" (p. 187). Those elements, together with how this study addressed them, are detailed in the following:

1. *Better defense against selection bias, through either randomized group assignment or pre-testing.*

While experiments that assign subject participation in either the experimental or the control group at random are particularly useful for establishing causality (Schneider, Carnoy, Kilpatrick, Schmidt, & Shavelson, 2007), the question of practical interest is ultimately whether or not, ceteris paribus, those inclined to study online learn as much as those inclined to study in traditional classrooms. Assigning participants randomly to the DE and CI groups may increase the validity of the empirical answer to a question that is not especially interesting or even sensible to ask, particularly in an age where exclusively face-to-face, exclusive online and hybrid programs all abound: how do those who may not otherwise ever be inclined to study online compare in terms of achievement to those who may not otherwise every be inclined to study face-to-face? Randomized assignment, thus, while useful for statistical control, may run contrary to natural student inclination and is undesirable as such. This study sought to split the difference by including an analysis limited to those students not taking more than 80% of their courses in one format

or the other, approximating a within subjects design, since the same students are taking both DE and CI courses (albeit to differing degrees).

2. Eliminating instrument bias and improving reliability.

Using published measures of student performance improves reliability and eliminates the potential for DE-CI bias that results from local instrument design, administration, and/or evaluation common in the DE-CI comparative literature. There is evidence that the consequences of instrument bias are not insignificant. Bernard et al. (2004) found in their meta-analysis that studies with researcher-made assessment instruments had an average effect size that favored DE, while those created by the researcher who was also teaching the courses in question had an average effect size favoring the classroom. Studies where the instrument was teacher-made but where the researcher was not involved in the teaching under investigation had an average effect size favoring DE. While the fact that none of these effects was significantly non-zero may temper the import of these distinctions, it does reinforce the potential for locally designed instruments to introduce bias into DE-vs.-CI studies.

This study obviated this concern by avoiding locally designed instruments outright. The MFT-B has a clearly delineated, well known set of expected outcomes, subject areas and question types, but cannot be influenced by any interested party to the research. Widely administered standardized tests like the MFT also provide published reliability statistics.

3. Improved control of operational factors.

In this study, the courses have the same syllabus, text, learning outcomes, assignments, and duration. All of the full-time instructors and many of the adjunct

instructors teach regularly both DE and CI classes. Class size is slightly higher on average in DE.

Other methodological desiderata for DE-CI studies beyond those explicitly cited in Bernard, Abrami, Lou, and Borokhovski (2004) include the following:

1. Controlling for the potential influence of additional possible confounding variables, including the following:
    - Student demographics, including age and gender: Controlled for statistically in this study.
    - Student ability: Given the absence of incoming student SAT or other standardized test data, this issue posed a challenge in this study. Overall GPA or general education GPA, however, may be a reasonable proxy and both were employed here.
    - Attrition rates, particularly given the evidence in the literature that satisfaction and persistence are higher in CI: This study, retrospective in nature, involved only those who have graduated, obviating the impact of any DE-CI retention differential. One issue that remains is the potential for a "survival of the fittest" bias. If persisting as a DE student is markedly more difficult, then one could posit that among those students studying entirely or predominantly at a distance, those who make it to graduation are for that fact alone likely to be more successful. This issue is not trivial. Fortunately, the issue is an empirical one, and the correlation between percentage of credits taken online and expected persistence will be estimated statistically using logistical regression.

- Instructional activities and methods: DE-CI equivalency here is more problematic, not because of the nature of this research but rather because techniques and methods appropriate for the classroom may not be ideal for online and vice-versa. Indeed, in her widely cited manual of online teaching technique, Salmon (2004) calls for an entirely new set of principles and techniques for DE. Nevertheless, the consistency in approach from DE to CI in this study was high in general (indeed, perhaps even to a fault vis-à-vis the recommendations Salmon makes). At the institution used for this study, classes were developed in DE and CI formats by the same instructor and at the same time, and inter-format uniformity is a design objective. Differences in activities and methods are generally but not exclusively a matter of degree.

2. Experimenter bias if instruments are human-graded and evaluators aren't blind to group membership: MFT-B scores were beyond the influence of the experimenter here, so this is not a concern.
3. Individual differences resulting from the fact that some students may have learning styles more amenable to DE than others and vice versa: Allowing students to pick DE or CI on a course-by-course basis allows them to do so in accordance with their own learning styles. Constraining the study to those students who have taken significant numbers of both DE and CI courses helps mitigate this issue as well. If those who disproportionately select DE over CI are indistinguishable in terms of other variables of interest and do significantly worse or better than their CI-favoring counterparts, then surely this is a noteworthy result, learning style differences notwithstanding.

Conclusions from the DE-CI Comparison Literature

In spite of literally thousands of empirical studies comparing DE to CI in terms of student outcomes (U.S. Department of Education Office of Planning Evaluation and Policy Development, 2010), as well as the many meta-analyses reported herein, and indeed even at least one related meta-analysis of the meta-analyses (Tamim et al., 2011), there is no clear consensus regarding whether DE is better than, worse than, or comparable to CI in terms of learning outcomes in the literature. As evidenced in the meta-analytic literature cited above, there are ample individual studies reporting DE to be worse than, better than, and (most frequently) roughly equivalent to CI in terms of student achievement operationalized in a wide variety of ways. Results reported in the meta-analytic literature follow a similar pattern. So, with a large majority of studies and meta-studies reporting no significant DE-CI differences and the remainder roughly divided between CI's and DE's being preferable respectively, should one conclude confidently that on the whole Russell (1999) was correct, and that there is no significant difference between DE and CI?

Perhaps not, for the reasons discussed as follows. Regarding the primary DE-CI research, methods to arrive at this evidence are typically uneven at best and often "woefully inadequate" (Bernard et al., 2009, p. 1125). Regarding the second-order or meta-analysis, it is grounded on that very research, and thus similarly suspect. Those meta-analysts seeking to impose high methodological standards are often stymied by primary empirical research reporting is also often inadequate, sometimes preventing methodological integrity from even being assessed. In other cases, results reporting is inadequate to afford subsequent second-order quantitative analysis.

Further, the meta-analyses reviewed herein were characterized by effect-size results much more heterogeneous than would be expected by chance. The variation in results from individual studies– even in those meta-analyses that scrupulously screened candidate primary studies – is sufficiently high that it is hard not to conclude that if these studies are indeed measuring anything meaningful, they are almost certainly not measuring the same thing one study to the next. Given the decades-plus periods they often span (U.S. Department of Education Office of Planning Evaluation and Policy Development, 2010, for example, covers studies over the period 1996 - 2008), this seems unsurprising: The state of DE practice may have begun to converge, but it was surely wildly different in the late 1990s than it was in 2008. One of the other sources of the heterogeneities that indicate studies measuring different things is the fact that the studies indeed were measuring different things. While many of the studies sought to measure student learning outcomes, they operationalized those outcomes in very different and often parochial, homegrown ways, many of which leave the door open for teacher and/or researcher bias.

In sum, more rigorous methodology and more transparent reporting among primary DE-CI studies such as the one herein stand to add to the literature. Using a measure of student learning that is standardized and inter-institutional – such as that discussed in the following section – affords the opportunity for replication studies and more useful systematic review as well.

Second Main Thread: The Major Field Test in Business and its Correlates

Given that the main research question at hand is the importance of traditional classroom contact to student learning achievement, the preceding review of the DE-to-CI comparative literature is of course central. Equally important given that the mechanism used here for measuring that achievement is a standardized objective test is the literature that investigates such tests, their validity and reliability, and the potential explanatory factors associated with student performance on them. This literature – discussed in the following – while not as extensive as the DE-to-CI comparative literature discussed above, is also robust, and no less useful for informing the research herein.

Predicting Standardized Test Scores

A large and longstanding literature on both the utility and the problems associated with using nationally administered standardized tests exists, examining, for example, the predictive power such scores have for college GPA (e.g. Geiser & Studley, 2002) , or investigating the ways in which that power might vary across various groups (e.g. Zwick & Sklar, 2005) . Those pursuits, however, are qualitatively different than this work and do not stand to inform this research directly. While efforts to explicate the ability of standardized tests to *predict* other outcomes are widespread, research interest in such tests themselves as *ends*, to be predicted by other variables, has been narrower, although it appears to be growing recently (for example, for the MFT-B alone, Bagamery, Lasik, & Nixon, 2005; Bycio & Allen, 2007; Contreras, Badua, Chen, & Adrian, 2011; and Mason, Coleman, Steagall, Gallo, & Fabritius, 2010 provide recent analyses of score determinants). That growth likely is explained by the elevated role of standardized tests created by increased emphasis on outcomes for accountability and accreditation

expectations (see, for example, Erwin, 2005). Thus, practitioner and administrator interest in research such as that presented here that examines standardized test scores as ends themselves is likely to grow. This fact is by no means universally celebrated.

Learning Assessment and the Increasing Call for "Accountability"

Interest in assessing student learning has grown dramatically in the past decades among institutions of higher education as well as the accrediting and policymaking bodies that oversee them (Suskie, 2009). In 1987 then Secretary of Education William Bennett called upon accrediting organizations to make the assessment of learning central to the accreditation process ("Secretary's procedures and criteria for recognition of accrediting agencies," 1987). Stakeholders have gone from mandating input and resource measures to output measures such as retention and graduation to desiring direct evidence of the knowledge and abilities students acquire at college. Erwin (2005, p. 127) explains the basis for mounting pressure: "Unless mandated by states, the use of common test data is unlikely across all institutions. But without comparable data about student learning, it is impossible to draw defensible conclusions about student learning for a state's system of higher education." Thus, accrediting bodies may feel pressure if not to mandate the use of standardized data appropriate to inter-institutional comparison then at least to encourage it.

In spite of any backlash against the threat of "No Child Left Behind" style accountability's making entreaties into higher education brought about by Spellings' (2006) "A Test of Leadership" – or, indeed, perhaps because of it –enthusiasm for assessing learning abounds in the academy and among its accreditors. Ewell (2007, p. 11) calls accreditors "the major external driver of assessment in for the past decade".

Regarding business education specifically, major changes in the Association to Advance Collegiate Schools of Business accrediting standards occurred in 2003, adding a set of standards mandating "assurance of learning" (Thompson, 2004). These standards required explicit statements of learning goals and direct evidence of student achievement of those goals. This change jumpstarted business programs' interest in objective achievement data such as that provided by the MFT-B (Martell, 2007).

Standardized Tests as a Means of Outcomes Assessment

The current accreditation climate demands that institutions engage in student learning assessment. For measures of learning assessment to be put to larger, accountability-related ends, they must be comparable across organizations. Because accrediting organizations – if for no other reason than in the name of self-preservation alone – seek to be the means by which broader accountability is realized, accreditors are interested in assessments that lend readily inter-organizational comparability. Standardized tests are particularly well suited to the accountability movement's desire for broad, objective, normed, "report-card" data and thus accreditors' and their member institutions' interest in assessment via standardized testing is growing as well. The Middle States Commission on Higher Education, for example, has a published statement on the role of published tests in assessment (Suskie, n.d.). Shaftel and Shaftel (2007) go so far as to liken the assessment component of the Association to Advance Collegiate Schools of Business (AACSB) to No Child Left Behind itself.

Institutions increasingly see tests such as ETS' Major Field Test in Business, administered to seniors as an exit examination, as useful in meeting the assessment expectations of specialized accrediting bodies such as the AACSB and the Accreditation

Council for Business Schools and Programs (ACBSP) in business and well as regional accrediting bodies (Allen & Bycio, 1997; Black & Duhon, 2003; Contreras et al., 2011; Dolinsky & Kelley, 2010; Mason et al., 2010; Mirchandani et al., 2001; Rook & Tanyel, 2009). Indeed, Mason et al. (2010, p. 71) call such exit exams the "currency of choice for both institutions and accrediting bodies seeking to demonstrate student learning".

Arguments in Favor of Assessment via Standardized Tests like the MFT-B

Major arguments in favor of the use of objective standardized testing beyond its utility in cross-institutional comparison include its relative economic efficiency, in that the tests are relatively efficient to administer and very inexpensive to score (although Mason et al. (2010) argue against the MFT-B on economic grounds, citing its "unfavorable" cost-benefit performance). Shaftel and Shaftel (2007) note the advantages in terms of reliability and validity afforded by deep question pools, broad content coverage, the involvement of teams of psychometricians and other features afforded by the deep pockets and broad reach of organizations such as ETS. Karthanos (1991) compares the MFT-B favorably to similar alternative tests.

Arguments Against the MFT-B

The MFT-B is a multiple choice examination, and while Educational Testing Service (2011) asserts that the MFT goes "beyond the measurement of factual knowledge", it requires students to recognize a correct answer from among a set of incorrect, a task susceptible to claims of inauthenticity relative to learning assessment alternatives that focus on performance assessments. Note, though, that while this lack of authentic performance assessment may be fair game for the MFT, it is not a blanket truth for all standardized tests. The Collegiate Learning Assessment discussed below is in

particular a standardized test grounded in performance assessment and devoid of objective questions.

Standardized tests like the MFT-B are typically low stakes for the students taking them – their own performance on such tests is normally of little consequence to them personally. This is by design; the tests are intended primarily not to reflect the accomplishments of the individual students but rather those of their respective institutions. On the other hand, if students remain unmotivated to try their best on such exams, those exams' validity as measures of student learning could be significantly compromised. Indeed, in an examination of the factors associated with scores on the Collegiate Learning Assessment (CLA) (Klein, Benjamin, Shavelson, & Bolus, 2007), a standardized instrument measuring writing and critical thinking, Hosch (2010) found that the amount of time a student spends on the test is strongly correlated with his or her expected score. If the amount of time spent on the test predicts score, one could readily argue that motivation is being measured at least as much as student learning. While Hosch's analysis was limited to the CLA, there is no reason to think that the motivation problem is not spread throughout similar low stakes testing situations. Allen and Bycio (1997) found that students who received the incentive of extra course credit dependent on their MFT score performed significantly better than those who did not.

Allen (2003) voices concern with standardized tests' inexact alignment with the specific learning objectives of individual academic programs. Bycio and Allen (2007, p. 196) are plainspoken on the matter: "…it is difficult to write a set of multiple choice questions that adequately reflects the business curriculum across a wide range of universities". "Teaching to the test", whereby instructors focus more narrowly on those

skills known to be consequential on the test to the exclusion of other, potentially more valuable material, is also cause for concern (Firestone, Schorr, & Monfils, 2004).

Mason et al. (2010) argue against MFT-B's use not by calling the test's reliability or validity into question but rather its efficiency, suggesting that the cost of administering and analyzing the exam is unjustifiable. To offer evidence of this, they provide an empirical model showing that major GPA and entering SAT scores, together with a few demographic factors, go a long way to explain performance on the MFT-B. They conclude that because MFT-B scores are highly predictable, the exam is needless relative to its costs. The goal of the MFT-B exam is assessment of the mastery of specific business content, however, and even though one may be able to predict who might be most successful, that does not mean there is no utility in objectively identifying which – as well as understanding why –students have actually mastered the desired content. (Not to mention the double bind arising from the fact that if the MFT-B did not correlate well with major GPA, it would stand susceptible to validity criticism.)

The MFT and Predictive Modeling

Many of the earlier studies investigating the factors that correlated with MFT scores (Allen & Bycio, 1997; Frazier & Edmonds, 2002; Karathanos, 1991; Mirchandani et al., 2001; Stoloff & Feeney, 2002; Szafran, 1996) focus primarily on exploring the criterion validity of the MFT test itself. Perhaps because over time the criterion validity of the MFT, if it did not come to be taken for granted outright, was seen as a moot point given its increasingly widespread acceptance and use, later research tends to focus on student performance and its determinants, often in the name of "documenting and explaining MFT results" (Contreras et al., 2011). If one can gain favor with accreditors

by administering the MFT-B, its criterion validity is irrelevant as a practical matter. Understanding its determinants, particularly those in the curriculum, so that one can improve student performance over time, on the other hand, becomes paramount. And indeed, case studies of successful adoption of the MFT-B into a department's assessment procedures and its use in curriculum improvement also appear (e.g. Bush, Duncan, Sexton, & West, 2008).

As Mirchandani et al. (2001) note, for virtually any standardized test, the best predictor of performance is some other standardized test. So too for the MFT-B: Mason et al. (2010) note that in their study of factors correlative with MFT-B scores that SAT score had a parameter estimate that was larger relative to its standard error than that of major GPA, and that increasing SAT scores is more likely to boost MFT-B scores than improving major GPA. But while the SAT is as much as anything a measure of general intelligence or test-taking aptitude (Frey & Detterman, 2004), the MFT-B measures business content mastery; certainly many individuals scoring well on the SAT and other measures of general aptitude, but with no training in business can be expected to do relatively poorly on the MFT-B. Admittedly, to the extent that SAT scores can predict MFT-B performance, the MFT stands to be as much a measure of the quality of a program's student body in terms of innate academic aptitude as it does a measure of the quality of a program itself. That fact does not preclude, however, two programs with equivalent incoming student SAT scores from producing students that perform qualitatively differently on the MFT-B at the end of their studies, nor does it preclude in such an event an interest in understanding why.

While the SAT's predictive power may be seen as problematic (as discussed in the following section), one could hardly blame the Educational Testing Service if it saw the correlation between major GPA and MFT score as an entirely desirable result. In the absence of such a relationship the MFT is exposed to criticism based on its lack of construct validity and in the presence of such a relationship it stands apparently susceptible to charges of redundancy and irrelevance.

Results from MFT-B Predictive Studies

Allen and Bycio (1997) found the single best predictor of MFT-B scores among both accounting majors specifically and all business majors generally was student GPA in core business classes, with verbal SAT scores a very close second. Accounting majors performed substantially better on the MFT than the other business majors analyzed, although major comparisons were conducted via ANOVA and did not control the variables found to be significant in the regression model. While accounting majors performed better on the test, they also had higher average SATs than the other majors. They found no significant difference for genders. Stepwise regression led to the inclusion of business GPA, SAT-V and SAT-M as significant variables and an adjusted R^2 of .55. Even with just these three variables, significant multicollinearity is an ex ante threat. Given the tendency of the independent variables to be correlated among one another, separating their effects meaningfully can be difficult, and estimated coefficients can be quite sensitive to small fluctuations in observed data. While that fact may remain, note that Bagamery, Lasik, and Nixon (2005, p. 58) examined grades course-by-course and found an "interesting" lack of multicollinearity among individual course grades. The

disaggregation of grade measures such as major GPA, business core GPA or overall GPA can be expected to further increase the potential for multicollinearity, however.

Bagamery, Lasik and Nixon (2005) took a novel approach to both the potential multicollinearity among various grade measures and the choice of a factor analysis to reduce the dimensionality of 14 course grades and found four significant factors, which they then used as independent variables in their regression model. Names for factors are subjective but they identified theirs as general, quantitative, accounting and management GPA factors. Significant variables beyond the grading factors included gender; age, transfer status, major, and test site (on-campus vs. off) were not significant. They also found a set of significant interaction terms between gender and grade factors that showed that the impact of increased GPA is less consequential in MFT-B terms for men than it is for women. Adjusted R^2s for their models were just under .5.

Black and Duhon (2003) found that business core GPA and composite ACT score were significantly, positively correlated with MFT-B results, and that men have a significant, 3.8 point score advantage all else equal over women. They divided majors into management versus non-management business majors and found management majors to be at a 3.6 point disadvantage relative to other business majors. They report an adjusted R^2 that is relatively high among similar studies at .58. Correlation between business core GPA and composite ACT scores was a reasonably modest .54, so multicollinearity was not seen as a major threat.

Bycio and Allen (2007), revisiting their previous research on the topic (Allen & Bycio, 1997), found results that mostly confirmed their previous study, with a few exceptions. Using a Likert-scale pre-test survey, they measured student motivations

explicitly. Using step-wise regression they found that only three variables – business core GPA, SAT-verbal and that motivation scale were required to account for 67% of the MFT variation. Once again, the ability of general knowledge and skill in the core business domains (as measured by core business GPA) was roughly equivalent to that of general intelligence and test-taking ability (as measured by SAT-V). Once again Bycio and Allen found no evidence of gender differences but, contrary to their previous work, no evidence of significant differences by major this time.

Contreras et al. (2011) specified a model of MFT-B score that includes dummy variables for major, age, gender, race (white versus black versus other only), average undergraduate GPA and composite ACT score. The found that compared to the general business majors, accounting, management and marketing majors were less likely to score well on the MFT. GPA and ACT were both found to significantly, positively correlate to MFT score. Older age was positive and significant, as was being male. Both the dummy variable for white and for black were positive but not significant. The authors then re-estimate the model for each of the major cohorts separately. With the exception of ACT score, whose coefficient remained similar and consistently significant across the major-specific models, significance, size and even the sign of parameters varied from one major sample to the next.

Dolinsky and Kelley (2010) analyzed the factors impacting scores in the MFT in psychology at a small Mid-Atlantic teaching college. Stepwise regression led to the same set of best predictors as studies discussed above, plus an additional factor not explicitly modeled in previous studies: Verbal SAT score was the best predictor, followed by major GPA and number of major credits earned upon taking the test. Dolinsky and

Kelley report that psychology faculty members were disappointed, given that the students in their program typically had SAT scores around the 50th percentile, that the first batch of student results showed MFT scores barely above the 25th percentile. Given the findings of Allen and Bycio (1997) and Bycio and Allen (2007) regarding the relationship between student motivation and scores, the fact that Dolinksy and Kelley's students knew that the test would not affect their grades and was to be used strictly to assess the program itself may go some distance to explain this underperformance. Frazier and Edmonds (2002) regressed MFT in psychology against individual student grades in each of the core psychology major courses and found that only the research methods course grades was significant, with a significant portion of the variation accounted for (R^2 = .701).

Stoloff and Feeney (2002) also examine determinants of scores in the MFT in psychology. They found a strong correlation between MFT-P scores and verbal SAT scores as well as GPA. A small but significant correlation was seen between MFT scores and the number of major courses completed. A subset of four individual courses in the major was associated with gains in the test.

Mason et al. (2010) provide a dataset that is larger than the other studies reported herein (1,411 students who took the MFT-B between 2005 and 2007, yielding a usable sample of 873). They estimated using OLS a model involving business class GPA, non-business GPA, age, gender, transfer status, a set of race/ethnicity dummy variables, cohort dummies for date students sat for the examination and interaction terms between cohort and business GPA. In keeping with other studies, business GPA and SAT were strongly significant. The only racial category significant at .05 was Asian, which fell

short of predicted performance for the referent category by just over two points. Age was highly significant and positive, as was being male. A 1.0 point increase in business GPA was associated with an 8 point increase in MFT score, while being male led to an expected increase of almost 5 points. Such a strong gender bias is noteworthy here, particularly given the absence of gender bias seen in other, similar studies (e.g. Allen & Bycio, 1997; Bycio & Allen, 2007). A subset of major categories was significant as well, with the only majors not outperforming the referent management majors being marketing and transportation/logistics, a highly specialized and somewhat marginal (relative to MFT-B test coverage) major. Results were consistent across testing periods. Goodness of fit was similar to other authors, with an adjusted R^2 of .565.

Mirchandani et al. (2001) examined grades in 13 individual business courses and conducted a factor analysis, finding that 11 of the 13 loaded on three factors, which they interpreted as quantitative courses, qualitative courses, and economics courses. They developed for students native to their university (i.e. non-transfer students) a regression model of MFT-B scores with independent variables including the three course GPA factors and SAT scores. While, consistent with other research, SAT scores were significant at the .01 level, none of the GPA factors were significant. For transfer students, for whom SAT scores were not available, they estimated a model using the GPA factors, supplemented with transfer GPA and found only the quantitative GPA factor to be significant. They re-estimated their model for male and female subsets and found a few differences. Among transfer students for whom SAT scores were lacking, the quantitative GPA factor was the only coefficient significant at .05, while for the female transfer students the economics factor was the only coefficient significant at that

level. Goodness of fit was low relative to similar research. Adjusted R^2 for the native student model was .371 while for the transfers it was .168.

Rook and Tanyel (2009) modified the prevalent approach by conducting a matched pair design whereby the same students were given the MFT during their sophomore year and then again during their senior year. The matched-pair design obviates the need for modeling SAT and may enable gains in scores from the first testing period to the second to be more directly ascribed to student learning (although within-group carryover effect may be a factor as well). Using percentile change as the dependent variable, core GPA was significant at the .001 level. A one-point improvement in GPA was associated with a 15 point percentile improvement in MFT-B. With an R^2 of .155, Rook and Tanyel's model explains relatively little of the variation in MFT-B scores.

Summary of the Empirical Research on MFT Prediction

Patterns of consensus as well as of open questions emerge from a review of the literature on factors contributing to MFT results. First, it appears clear that measures of general intelligence and/or test-taking skill such as SAT or ACT exam scores can be expected to strongly predict MFT scores. While this may not be uniformly encouraging news for the validity of the test as a measure of student mastery of core business skills and knowledge, it is a consistent finding in the standardized test literature. Similarly, GPA in the core discipline is a strong predictor of MFT success, although generally not quite as strong as SAT.

The strong correlation of SAT/ACT and MFT-B scores reported in the literature, because they do not capture the value programs add to their students' content matter

learning, present those seeking to understand the factors controllable by institutions and program a challenge, particularly to the extent that they stand to cloud through multicollinearity the roles of variables such as major, and core business GPA that more strongly represent content mastery. Rook and Tanyel (2009) address this through a matched-pair design, modeling the improvement test takers experience in the 4-year MFT from their sophomore to their senior year. While it is not the approach Rook and Tanyel took, the MFT in business is offered at the associate, bachelor's and MBA levels, so the potential exists to use, for example, a matched pair design to model improvements from the associate test to the bachelor's, or to use the associate test as an independent variable in modeling the bachelor's score. Because such tests are independent of one another, carryover effects are less an issue in such a design as well.

Regarding demographic variables, there is no clear consensus regarding gender or racial bias; some saw significant differences, particularly for men versus women, but others did not. Age, where modeled, tended to be positively correlated with MFT score. The role of academic major is unclear, and appears idiosyncratic per school studied. There is no evidence that being in a particular major puts one at a significant disadvantage, which is good news for a test that purports to assess core business skills common to all business majors.

Grades in individual classes and their relationship to MFT-B scores was a theme that reappeared in the literature. Particularly for those institutions looking for mechanisms to improve MFT scores through curricular interventions, this thread would be of particular interest. Authors typically reduced the dimensionality of the per-course grade data via factor analysis before using it to fit the MFT result data. While SAT

verbal was consistently found to be a better predictor of MFT score (in both business and psychology), when it comes to grades, the course grade factor associated with quantitative skills tended to have better explanatory power. Goodness of fit statistics for MFT models range from R^2 of .15 to .7. (It was not always clear whether the reported coefficient of determinations were adjusted for the number of explanatory terms or not).

Research on factors predicting MFT-B performance reviewed above has been limited to single-institution datasets. As Contreras et al. (2011) noted, research to broaden the variety of institutions studied would be helpful. A cross-institutional study would be ideal, although inter-institutional sharing of this sort of student performance data may run afoul of privacy concerns and seems unlikely.

None of the literature on MFT-B identified herein addressed the question of online versus classroom education. Given that the research into explanatory models of MFT scores has its roots primarily in investigating the validity of the test itself, and later in providing insight on to how departments might improve their scores, this is unsurprising. Nevertheless, extending the approach taken in the extant MFT models to include online versus classroom instruction seems a worthwhile foundation for further understanding the effectiveness of modern distance education relative to that occurring in the classroom. Such an effort is particularly attractive as accrediting bodies' interest in scrutinizing online education and its assessment continues to grow.

Conclusions from the Literature Review

The impact on student learning outcomes of studying at a distance as opposed to via classroom instruction is, in spite of a huge body of empirical work, still an open

question. This fact is likely due in no small part not only to inadequate methodological rigor on the part of the primary studies but also to undisciplined and incomplete reporting of method and results that stymies principled, systematic meta-analysis. The meta-analyses that have been conducted report widely heterogeneous results, implying that they are seeking to synthesize underlying studies that are measuring different things; given that these studies often employ locally developed measures of learning outcomes, they indeed often literally are.

Consistent with the general trend of accountability in higher education, Business programs are increasingly employing Educational Testing Service's Major Field Test in Business to provide the sort of objective measure of their students' content knowledge increasingly favored by accreditors. As such, the MFT-B's correlates are increasingly well reported in the business literature. The impact of studying all or partly at a distance on MFT-B scores does not yet appear as a factor in this literature.

The DE-CI comparative literature historically has characterized the decision to study at a distance as an all-or-none-proposition, investigating its consequence in a between-groups fashion. Students, however, are increasingly mixing and matching DE and CI courses, as well as participating in blended courses and thus studying at a distance is better understood as a question of degree. When categorizing specific educational experiences as either DE or CI, the literature has typically stratified according to the technology haves and have-nots, categorizing any course with a significant online component – irrespective of the presence of significant classroom interaction – as DE. As traditional classroom instruction is increasingly supplemented with the tools of DE as

a matter of course, the more vital question is the impact of the presence or absence of classroom interaction.

Using the MFT-B as the measure of student learning outcome, this study modeled DE as a proportion of total credits conferred, in a program comprised of hybrid courses with DE tools supplemented with significant classroom interactions versus those occurring using the same tools, but exclusively asynchronously and at a distance. No similar such study appeared in the literature. This study may enable a better understanding of the consequences of the decision to forgo classroom interaction, in a context that more accurately reflects the way in which students currently undertake that decision than has appeared in the literature heretofore.

CHAPTER 3
METHODOLOGY AND PROCEDURE

Introduction

This study sought to understand the impact of the proportion of coursework that students in a hybrid undergraduate business program at pseudonymous Urban College completed at a distance had on their performance on the MFT-B examination that they took at the end of their studies. The dependent variable in this study was the sample students' MFT-B exam performance. The dependent variable was modeled as a function of the percentage of courses completed in the DE format, together with additional factors suspected to impact MFT-B test performance, which include student demographic variables (age, gender, and race), major, the proportion of credits transferred from another institution, and cumulative GPA. This study sought also to estimate the impact of study at a distance in the same context on likelihood of graduation.

Role of the Researcher and Ethical Issues

I am currently an Associate Professor of Information Technology and an Assistant Dean of General Education and Information Technology at the institution under examination. I do not teach in the business program under study here, nor do I teach any classes normally taken by business students. As an educator regularly teaching both traditional classroom-based and entirely asynchronous DE sections of the same courses, I have found myself struggling to achieve the same gains online that come readily in the physical classroom. Perhaps this is because I normally teach technical material, often to students – whether online or in the classroom – who are relatively unprepared to take such material on, or perhaps because the material – computer programming in particular

– is "hands-on" in a way that resists teaching at a distance. Alternately, my teaching style may be idiosyncratically incompatible with instruction at a distance, or perhaps I am less able to appreciate the gains remote students achieve. In any case, I've always marveled at the "no significant difference" DE-CI studies in light of my own teaching experience, and was skeptical at the outset of the prospect that DE might lead to improved learning outcomes.

In my capacity as an administrator, however, I may be inclined to look upon DE more favorably because it is a source of over half of Urban College's enrollment and is growing strongly. I have difficulty imagining a prosperous future for Urban College that doesn't involve DE figuring centrally.

Irrespective of the net result of any potential biases I may have as the researcher, my role, together with the fact that the data analyzed here have been collected retrospectively and involve record of events that have already occurred and were beyond my influence then or now, should limit the potential for consequence arising from said bias.

Ethical issues in the study were limited to those related to the potential harm resulting from disclosure of subjects' private personal information. Given that the information collected involves course grades and cumulative GPAs, the potential is not insignificant. All data were de-identified at the time they were retrieved from the student information system, however, mitigating the potential for disclosure.

Population and Sample

The study involved students at "Urban College", a small, private, specialized, not-for-profit college providing career-focused associate and bachelor's degrees, primarily to "nontraditional" students. While a clear and consistent definition of what constitutes a nontraditional student remains elusive (National Center for Education Statistics), in Bean and Metzner's (1985) review of the literature, age was one of the most common defining characteristics. The mean age of Urban College students was 34 and the mean age of the 817-student sample used in the study was 36 years; 83 of the 817 were under 25, a common cutoff age for defining traditional student age. Age, however, is largely a proxy for other more consequential defining characteristics. Nontraditional students typically work full-time, are financially independent, have family responsibilities, commute to school, and study part time. Thirty-six percent of Urban's students studied full-time in fall 2010 ("National center for education statistics college navigator," 2011).

Jones and Watson (1990) draw a distinction between "high risk" students and nontraditional students, noting that, while there is often significant overlap between the two categories, they are not equivalent. They identify minority status, academic disadvantage and poverty as indicators of high-risk status. Urban College also has a large number of high-risk students by this measure, as well as a many first generation college students.

The study sample included students pursuing a bachelor's degree in one of Urban College's business majors – accounting, business law, human resource management, general management, marketing, real estate management, small business management, or technology management – who took as part of their capstone business class the MFT-B

between May, 2004 and August, 2009 (the dates for which test score records were available in the Urban student information system).

Data Collection Procedures

While similar data are routinely collected by Urban's Institutional Research department, particularly for accreditation purposes, to enable full control over the data obtained, I retrieved these data from Urban's Jenzabar student information system personally, via direct SQL query. My access to such data was within system permissions associated with my current role and position. All data analyzed in this study were obtained in this way. Separate queries were required to identify students sitting for the MFT-B exam, for their relevant demographic information, for their GPA, and for their course histories. With the exception of transcript information detailing which courses were taken in which format, when and what grade was earned, the remaining data were fashioned into one unified data set with the individual student as the unit of measure. Data validity and accuracy concerns beyond those occurring at the time of original data entry are limited.

My colleagues in the business department were aware of my research at the outset and indeed their interest in the questions posed herein was among the motivations for conducting this work in the first place.

Data Analysis Procedures

The study's major research question involved assessing the extent to which the data provide evidence that students who complete a large portion of their studies at a distance perform on the MFT-B significantly differently than those who study primarily

in a mode involving regular classroom interaction, holding all other influential factors constant. Doing this involved developing a model of MFT-B scores as a function of proportion of coursework completed online together with the other predictor variables. This multiple regression model was estimated using ordinary least squares via SPSS/PASW for Windows version 19.

Several of the independent variables – race, gender, and major – are nominal in scale and thus require modeling using "dummy" variables. When setting up dummy variables, a nominal variable with *n* possible states results in *n-1* mutually exclusive binary categorization variables, with a value of one indicating a given observation's membership in the given category and an absence of the value of one in any of the *n-1* category variables indicating membership in the (arbitrary) default or reference category. Given that race is classified according to eight categories and that there are 9 business majors of interest, the presence of these multi-category dummy variables stands to reduce the model's degrees of freedom and thus its power considerably. Fortunately, with on the order of 800-900 observations, this concern is mitigated somewhat. These concerns can be further limited by using stepwise regression or by consolidating categories with limited membership.

Bagamery et al. (2005) found gender to be a significant variable in their predictive model of MFT-B scores in that their academic performance factors under-predicted MFT-B scores in their sample's male students. They conducted a factor analysis using individual course grades and identified four subject area GPA factors. Positing that the MFT consequence of one or more of those academic performance factors in the overarching predictive model might be influenced by gender, they created a series of

interaction variables between the academic performance variables and gender. They found the interaction between what they name the "general business" subject GPA factor and gender to be highly significant, with unit improvement in that factor's GPA increasing expected MFT-B disproportionately in male students.

To identify sets of courses whose GPA is particularly influential on MFT-B score, a course grade factor analysis a la Bagamery et al. (2005) was conducted. Similarly, while the majority of predictive variables appear a priori to be straightforwardly additive, interaction between gender and GPA and especially between gender or age or race/ethnicity and the influence of distance education proportion may be present. The impact of proportion of coursework completed online could easily be influenced by a student's gender, in a fashion similar to the interaction between GPA and gender reported by Bagamery et al. Given that the interactions involved the binary (for the purposes of this study) gender variable, rather than model a series of interaction terms, the study, in addition to estimating a model for all observations, estimated it for female students only and again for male students only and compare the results. Interaction between GPA and proportion of online study, however, was modeled explicitly because the influence of GPA on MFT-B could be disproportionately muted for those studying mostly online (or vice versa).

Issues Related to Operationalizing Proportion of Study Completed at a Distance

Credits study participants earned locally at Urban are known to have been earned in either the hybrid or the DE format. This is not the case, however, for credits transferred from other institutions, creating a challenge in determining how best to measure how much of a given transfer student's studies have been completed at a

distance. Using percentage of local credits earned online was attractive because it reflected students' inclination to take the DE format when it is offered. DE adoption at other institutions varies, however, and in general DE is likely to be less widely available elsewhere than it is at Urban (save perhaps for proprietary schools). Because inclination to study via DE is irrelevant when the option is unavailable, percentage of local credits likely overstates the total proportion of coursework completed via DE. Counting DE credits, or equivalently figuring the percentage based on the 120 credits required to graduate from the Urban business program, understates the extent of online study for any student transferring credit completed in the DE format. Neither approach is ideal. Models were estimated using both representations and the results' sensitivity to the choice analyzed. Outcomes are discussed in the following chapter.

CHAPTER 4
STUDY RESULTS

Characteristics of the Sample

The students of interest in this study were all Urban College business administration students who sat for the MFT-B test. The test is administered during the senior capstone class required of all business students and taken in the last semester prior to graduation. MFT results are entered by hand into the student information system by the Registrar's office after each test administration. Students studying at a significant distance from Urban arrange to take the MFT at another college or other authorized testing center; local students (including those studying mostly or entirely via DE) take the test at Urban. Records of 817 students having taken the MFT-B exam were available for analysis. For each of the 817 students, the demographic characteristics reported in Table 4.1 below were retrieved from the student information system, together with a record of the number of courses each student transferred to Urban, all courses taken locally, their respective grades, and whether the course was taken in the DE or hybrid format. These data were de-identified immediately upon extraction from the student information system and assigned a synthetic tracking number. Students in the sample sat for the MFT-B test on 20 occasions between October 2005 and August 2009.

The demographics of Urban College generally (12-Month IPEDS Enrollment for 2010, presented in Table 4.2 below) roughly match those of the study sample: the large majority of the sample students were female; the majority was African American, and their mean age at the time of testing was 36 years as compared to Urban's overall average

of 34 years. Slightly fewer than half of the credits earned locally were taken in the DE format. A full account of the descriptive statistics for the sample is as follows:

Table 4.1. Descriptive Statistics for MFT Student Sample

Variable	N	Minimum	Maximum	Mean	Std. Deviation
MFT-B Score	817	120	186	140.59	11.02
Transfer Credits	817	0	90	25	27.62
General Education GPA	813*	1.42	4.00	3.26	.5637
Major GPA	817	1.67	4.00	3.34	.4972
Local online credits as a percent of total local credits	817	0%	100%	48%	.3523
Local credits earned online	817	0	145	46	35.23
Age at time of MFT test	817	21	70	36	9
Gender = Male	155/817	n/a	n/a	19.0%	n/a
Major:					
Accounting	128/817	n/a	n/a	15.7%	n/a
Business Law	33/817	n/a	n/a	4.0%	n/a
Human Resources	13/817	n/a	n/a	1.6%	n/a
Management	579/817	n/a	n/a	70.9%	n/a
Marketing	43/817	n/a	n/a	5.3%	n/a
Other Business	21/817	n/a	n/a	2.6%	n/a
Ethnicity:					
African American	480/817	n/a	n/a	58.8%	n/a
Hispanic/Latino	36/817	n/a	n/a	4.4%	n/a
Caucasian	228/817	n/a	n/a	27.9%	n/a
Unknown	40/817	n/a	n/a	4.9%	n/a
Other	33/817	n/a	n/a	4.0%	n/a
Time to completion (normalized to 120 credits) in years	791**	1.83	45.09***	6.52	4.011

* Four students transferred all of their general education requirements from other institutions, precluding a General Education GPA calculation.

** Start dates for 56 students were unavailable in the SIS.

*** One student did indeed take 25 years to earn 67 credits, starting in 1982 and finishing in 2006; she did not study continuously during that span.

Table 4.2. Urban College Demographics

Category	Percent
Men	28.8%
Women	71.2%
African American	58.9%
Caucasian	28.3%
Hispanic/Latino	5.6%
Asian/Pacific Islander	1.1%
American Indian/Pacific Islander	0.6%
Race/ethnicity unknown	3.1%
Non-resident alien	2.4%

General education GPA is defined here as the student's average GPA for all courses taken that satisfy one or more of Urban's general education core requirements, irrespective of whether those credits were actually needed to satisfy those requirements. For example, if a student opted to take a literature or math course as a free elective, that course was included in the calculation of general education GPA for this study, in spite of the fact that the student did not need that class to satisfy his or her core requirements. The major GPA includes all of the courses that were required to satisfy the requirements of the student's chosen major. These courses include the core business classes required of all business majors plus those courses required of each specific major.

For categorization by major, any major accounting for less than 1.5% of the sample was collapsed together into the category labeled as "other" in Table 4.2 above. These majors included small business management, real estate management, and technology management. Similarly, for ethnicity, the non-citizen, Native American/Alaskan and Asian American categories, each representing less than 1.5% of the total sample each, were collapsed together in the catchall category "other" as well.

Research Question One

Research Question One is as follows: Holding all other known relevant variables such as GPA, age, race, gender, major, and semester standing constant, does the extent to which students in a bachelor's program in business study at a distance as opposed to in a hybrid format involving classroom interaction impact their performance on the Major Field Test in Business (MFT-B)? To address this question I modeled MFT-B score as a function of the proportion of studies completed in the DE format (the primary independent variable of interest for question one), transfer credits earned, general education (GE) GPA, major GPA, age at the time of testing, gender, major, ethnicity, and completion time (normalized to 120 credits), estimating the equation using ordinary least squares (OLS) regression. Coefficients for the resultant equations are reported in Table 4.3 below. Univariate, pair-wise correlations among the variables used in the equation can be found in the appendix. A constant is included in this model and subsequent variations because MFT-B scores have by design a floor of 120.

Table 4.3 Regression Coefficients from Modeling Online Study as the Percentage of Local Credits Completed Online

Model	Beta	Standard Error	Standardized Beta	t-score	P
(Constant)	109.104	3.335	N/A	32.714	< .001
Percentage of total local credits that were earned online	7.216	1.064	.229	6.784	< .001*
Transfer credits earned	.029	.013	.072	2.230	.026*
GE GPA	2.871	1.132	.143	2.535	.011*
Major GPA	5.465	1.316	.241	4.152	< .001*
Age at the time of testing (years)	.025	.039	.021	.654	.513
Gender = male	3.902	.853	.139	4.575	< .001*
Major = Accounting	.720	.954	.023	.755	.451
Major = Business Law	2.488	1.670	.045	1.489	.137
Major = Human Resources Management	-2.604	2.716	-.029	-.959	.338
Major = Marketing	.445	1.585	.009	.281	.779
Major = Other	6.018	2.104	.087	2.860	.004*
Ethnicity = African American	-3.862	1.599	-.171	-2.416	.016*
Ethnicity = Hispanic/Latino	-6.914	2.275	-.121	-3.039	.002*
Ethnicity = Caucasian	-.807	1.631	-.033	-.495	.621
Ethnicity = Other	-3.576	2.262	-.065	-1.581	.114
Normalized degree completion time(years)	.070	.086	.025	.810	.418

*p <. 05

The adjusted R^2 for this model is .311, relatively modest but in line with the similar regression models having MFT-B scores as the independent variable of Rook and Tanyel (2009), .155; Mirchandani et al. (2001), .371; or Mason et al. (2010), .565. At an

α of .05, eight variables in addition to the equation's constant are significant: Percentage credits earned online, number of transfer credits, general education GPA, Major GPA, gender, having a major of "other", and being African American or Latino.

Studying online is strongly associated with an increase in MFT-B performance ($p<.001$). Practical significance is not the same as statistical significance, but this result appears to be of practical significance as well. The predicted differential between the MFT performance expected of those studying entirely in hybrid courses meeting regularly and those studying entirely at a distance is 7.2 test points. As a rough indicator of practical significance, rounding to whole number, if the 7-point differential were split across the mean MFT-B score for the sample of 141, it provides a score range of 137 to 145, which, according to ETS (Educational Testing Service, 2011a) would move the institution's mean performance in 2011 from the 4th percentile for all domestic institutions administering the test to the 20th percentile. Such a change would likely be seen as of considerable practical significance to those responsible for the school's business program.[4]

By way of comparison to the impact of proportion of online study, a one point increase in GPA – from a 3.0 to a 4.0, for example, is expected to increase one's MFT-B score by 5.5 points, again an increase of undoubted practical significance, but less than the change expected from studying entirely via CI to entirely those held exclusively at a distance. Thus a 'B' student studying entirely online would be expected to outperform on the MFT-B a straight 'A' student studying entirely in the hybrid format.

[4] This is of course not to say that somehow moving all Urban students to study exclusively at a distance would *cause* such an improvement, but rather just to illustrate that the differential is sizable in practical terms.

Standardized betas provide a mechanism for comparing the impact of each coefficient relative to the others, their different, often entirely incommensurate scales and units of measure notwithstanding. Standardized coefficients allow this comparison by representing how much the dependent variable (MFT-B here) will change per unit standard deviation change in the independent variable. Major GPA had the largest standardized coefficient (.241), but percentage change in local credits taken online was a close second at .229. Standardized betas for all predictor variables can be found in Table 4.3. Note that coefficient standardization has its critics (e.g. King, 1986) and certain variables are as a practical matter far easier to change by a standard deviation than others; nevertheless, as a means to triangulate relative consequence, they are of some use here.

The coefficient for maleness was also positive and strongly significant, on a scale similar to that reported in some prior research on the correlates of MFT-B, e.g. Bagamery et al. (2005) and Mason et al. (2010), but not all: Allen and Bycio (1997) and Bycio and Allen (2007) report no evidence that gender is a factor in MFT-B performance. Relative to the referent ethnicity category of "unknown", two categories are significantly non-zero: African American and Hispanic/Latino. Both are lower relative to the unknown category, as is Caucasian, although not to a statistically significant degree at an α of .05. The impact of ethnicity reported in the existing literature on MFT-B correlates is inconsistent. The significance of the "other" business major category appears to be attributable to three individual students in the Real Estate Management program whose high scores amount to outliers when compared to that of the sample as a whole.

Issues Related to Representing the Extent of Study at a Distance

In the regression results reported above, credits earned online are calculated as a percentage of credits earned locally at Urban. However, average credits transferred into Urban by the students in the sample were fairly high at 25 and ranged widely with a standard deviation of over 27. Outside of the fact that for the most part community colleges have been slow to embrace distance education, it was not possible to determine whether credits transferred from another institution were earned online or in the classroom. This fact presented a challenge when representing the extent to which studies have occurred at a distance. In the model discussed above and summarized in Table 4.3 above, credits were calculated as a percentage of the total earned locally. This representation ignores the fact that students transfer credits to Urban to widely varying degrees (although those credits are themselves in fact explicitly modeled separately).

As an alternative, the absolute number of credits earned online was represented in the model, irrespective of the proportion that credit count represented of the credits for which the mode of study was known. Modeling distance education in this fashion results in the regression coefficient results reported in Table 4.4 below. Adjusted R^2 in this model is .301, a minimal reduction from the explanatory power of the percentage-based model. Neither the magnitude of the coefficients, either in absolute or relative terms, nor their significance differs notably from the model based on percentage representation. The impact on expected MFT-B from zero credits of study at a distance to a full 120 credits is here 7.08 points compared to the 7.2 points estimated in the distance education as a percentage model above.

Table 4.4. Regression Coefficients with study at a Distance Modeled as Count of Local Credits

Model	Beta	Std. Error	Standardized Beta	t-score	P
(Constant)	109.723	3.387	N/A	32.390	< .001
Count of local credits earned online	.059	.010	.187	5.798	< .001*
Transfer credits earned	.063	.013	.156	5.009	< .001*
GE GPA	3.043	1.140	.152	2.669	.008*
Major GPA	5.297	1.326	.234	3.995	< .001*
Age at the time of testing (years)	.018	.039	.015	.455	.649
Gender = male	3.827	.859	.136	4.454	< .001*
Major = Accounting	.573	.962	.018	.596	.552
Major = Business Law	2.400	1.689	.043	1.421	.156
Major = Human Resources Management	-3.073	2.754	-.034	-1.116	.265
Major = Marketing	.267	1.597	.005	.167	.867
Major = Other	6.623	2.113	.096	3.135	.002*
Ethnicity = African American	-4.287	1.607	-.190	-2.668	.008*
Ethnicity = Hispanic/Latino	-7.277	2.291	-.128	-3.177	.002*
Ethnicity = Caucasian	-.887	1.645	-.036	-.539	.590
Ethnicity = Other	-4.395	2.267	-.079	-1.939	.053
Normalized degree completion time(years)	.057	.087	.021	.657	.511

Because the two approaches to representing the extent of study at a distance do not differ substantively in terms of their predictive power or their specification, all subsequent modeling reported below will use the percentage-of-local-credits-obtained-at-a-distance approach.

Research Question One, Sub-question One

The models above assumed that the influence of each predictive variable is independent from, and thus additive to, all the others (as indeed do all linear regression

models that lack interaction terms). Thus, for example, in the model gender may have influenced expected MFT-B scores, and online study may have influenced MFT-B scores, but there is no mechanism for evaluating the potential influence gender might have had on *the extent to which* online study influences MFT-B performance. The first sub-question to research question one –"Do demographic characteristics such as age, gender or race change the way and extent to which proportion of online study impacts performance on the MFT-B?" – called for such a capacity.

There were several approaches to enabling the potential identification of such interactions. The first was to add interaction terms to the original model. The first term added was the gender variable (coded as 1 for male, 0 for female) multiplied by percent of online study. This term allowed the potential for the extent of online study to have an impact of a different magnitude for males than it did for females. If the interaction term was significantly positive and the online coefficient remained positive as well, then males are expected to be more sensitive to the positive impact on MFT-B of online study, and so forth. Adding an interaction term between gender and percent online, however, did not result in a significant beta at $\alpha = .05$ and thus there is no basis in these data for rejecting the hypothesis that the impact of online study is equal between genders.

Binary (for the purposes of this research at least) categorical variables such as gender provide a straightforward alternative to modeling explicit interaction terms; however, one may estimate the original model only for the subset of the sample where gender is female and then again where it is male. Of course this reduces the sample size significantly and thus the resultant model's statistical power, particularly in this case

where only 19% of the sample is male, but the sample size here was large enough to have made it worthwhile to estimate nonetheless.

Re-estimating the model reported in Table 4.3 for the 662 female subjects alone resulted in the coefficients reported in Table 4.5 below. The female only model's goodness of fit remains unchanged from the full sample model ($R^2 = .311$). Coefficient magnitude and significance changed little as well. The business law major coefficient became significantly higher than the referent management major at $\alpha = .05$ level and the reduction in expected test scores for membership in the African American ethnic group relative to that of the unknown ethnicity category fell below the established .05 significance level. Most importantly, the coefficient for study at a distance remained in the same 7-8 point range it has for each of the models reported previously, and it also remained strongly significant.

Table 4.5. Regression Coefficients, Female Observations Only

	B	Std. Error	Standardized Coefficients	t-score	P
(Constant)	109.994	3.502		31.406	< .001
Percent of Local Credits Earned at a Distance	7.593	1.130	.255	6.723	< .001*
Total Transfer Credits	.031	.014	.082	2.295	.022*
GE GPA	3.765	1.170	.198	3.218	.001*
Major GPA	4.377	1.352	.205	3.238	.001*
Age at the time of testing (years)	-.003	.041	-.002	-.065	.948
Major = Accounting	.183	.973	.006	.188	.851
Major = Business Law	3.707	1.733	.072	2.139	.033*
Major = Human Resources Management	-2.645	2.785	-.032	-.950	.343
Major = Marketing	2.108	1.680	.044	1.254	.210
Major = Other	4.812	2.185	.075	2.203	.028*
Ethnicity = African American	-3.296	1.739	-.154	-1.895	.059
Ethnicity = Hispanic/Latino	-7.304	2.543	-.126	-2.873	.004*
Ethnicity = Caucasian	-.014	1.805	-.001	-.008	.994
Ethnicity = Other	-4.352	2.491	-.080	-1.747	.081
Normalized degree completion time(years)	.081	.085	.033	.944	.345

Repeating the model for only the male subjects, however, resulted in a model in many ways quite different. Estimating a 15-variable equation with 155 observations, 151 of which have values for all 15 variables, resulted in a marked reduction in degrees of freedom and thus of statistical power. R^2 for the male-only model dropped to .241. An analysis of variance showed that, with an F of 4.201, the proportion of variance attributable to the regression is still strongly significant ($p < .001$) but dramatically reduced from previous models with several times more observations.

The coefficients for the male-only model are reported in Table 4.6 below. For the male-only model, the only coefficient significant at the .05 level was major GPA. The coefficient for study at a distance dropped from the 7-8 point range seen in the previous models to under 5, with a large increase in its standard error, resulting in a an significance well short of the .05 threshold, or even twice that. One is left to speculate whether this was because men at Urban College are less susceptible to the benefits of study at a distance in terms of expected increase in MFT-B score or if it was because of a relative dearth of male subjects. The lack of statistical significance in the gender – distance education term discussed above was subject to exactly the same speculation, and estimated (as expected) an MFT-B score impact for males almost exactly that of Table 4.6's equation (4.6 points there versus 4.7 here).

Table 4.6. Regression Coefficients, Male Participants Only

	B	Std. Error	Standardized Beta	t-score	P
(Constant)	103.548	9.318		11.113	< .001
Percent of Local Credits Earned at a Distance	4.670	2.895	.133	1.613	.109
Total Transfer Credits	.019	.038	.040	.501	.617
GE GPA	-1.124	3.406	-.050	-.330	.742
Major GPA	11.125	4.052	.432	2.745	.007*
Age at the time of testing (years)	.184	.111	.132	1.655	.100
Major = Accounting	4.284	2.992	.104	1.432	.154
Major = Business Law	-4.768	5.212	-.072	-.915	.362
Major = Human Resources Management	-3.153	8.295	-.028	-.380	.704
Major = Marketing	-5.282	4.195	-.091	-1.259	.210
Major = Other	9.793	6.296	.121	1.555	.122
Ethnicity = African American	-5.205	3.996	-.199	-1.302	.195
Ethnicity = Hispanic/Latino	-6.542	5.230	-.125	-1.251	.213
Ethnicity = Caucasian	-2.641	3.904	-.099	-.676	.500
Ethnicity = Other	-1.709	5.529	-.031	-.309	.758
Normalized degree completion time(years)	.210	.355	.047	.591	.556

The coefficient for general education GPA, while not statistically significant, was negative here – an unexpected result – and major GPA roughly doubled in magnitude from previous models. This result raised the alarm regarding multicollinearity between the two variables. Perfect multicollinearity – that is, a correlation of +/- 1 – makes estimating a model impossible; variables that are highly but stochastically correlated, however, do not preclude model estimation or reduce explanatory power, but, because their respective influence on the dependent variable relative to one another is difficult to disentangle one from the other, their coefficients are fragile, and subject to sizable

changes from small, potentially random disturbances in the values of the variables themselves.

I expected general education GPA and major GPA to have been positively correlated because they are both partially determined by some broader underlying latent academic competence. Indeed, the bivariate Pearson correlation between the two variables was .844. Pair-wise correlation above .8 is alone enough to raise multicollinearity warnings (Mansfield & Helms, 1982), and indeed, the Variance Inflation Factor (VIF) for General Education GPA was 3.65 and for Major GPA was 3.85, indicating in both cases a variance nearly twice as high as would be expected if the variables were uncorrelated (although falling short of the common heuristic cutoff of a VIF of 5 – 10 as an indication of problematic multicollinearity)[5]. In a situation similar to that of the relationship between gender and the influence of online study, it is possible that for males the influence of general education GPA was negligible and that of major GPA was much more consequential, but it is also quite likely that inadequate data, in this case along with significant multicollinearity were actually at issue.

Modeling overall cumulative GPA instead of both general education GPA and major GPA together skirts the GPA component multicollinearity problem altogether, but

[5] In addition, the categorical variables for race = African American and race = Caucasian displayed evidence of multicollinearity as well (VIF = 5.7 and 4.9 respectively). Because membership in these two categories together represents almost 90% of the total sample, the presence of both of these variables in the equation creates a stochastic approximation of the perfect multicollinearity of the so-called "dummy trap". This multicollinearity could be remedied by collapsing the 5 variable ethnicity/race classification scheme into white and nonwhite. Such a scheme would generalize racial identity to an undesirable degree and thus will not be presented here. Rather, note that estimates for the coefficients for racial/ethnic categories may be fragile and subject to large changes from relatively minor, potentially random changes in the underlying variable values. Note further that collinearity need not be of the zero order, pair-wise sort discussed here.

in practical terms it yields very nearly the same result as the model with problematic zero-order multicollinearity: a sense of the impact of GPA overall without confidence about the respective contributions of general education versus major courses respectively. The results from this single-GPA model are reported in Table 4.7 below. Compared to the original two-GPA model summarized in Table 4.3 above, the single-GPA had a slightly improved adjusted R^2, up from .311 in Table 4.3 to .347 here (partly as a result of the increase in degrees of freedom afforded by reducing the number of predictor variables by 1). The coefficients and their significance, however, remained remarkably stable between Table 4.3 and Table 4.7 and indeed, the collective impact of general education and major GPA in the two was fairly similar – 11.7 here versus 8.3 between general education and major GPA in Table 4.3. And indeed, with the exception of the African American and Caucasian racial/ethnic categories (again for the reason detailed above) all VIFs were well below 2 in the revised, single GPA equation.

Table 4.7. Regression Coefficients, Overall Cumulative GPA Model

Model	Beta	Std. Error	Standardized Beta	t-score	p
(Constant)	97.193	3.677		26.430	< .001
Percent of Local Credits Earned at a Distance	6.935	1.026	.220	6.757	< .001*
Cumulative GPA	11.666	.907	.424	12.868	< .001*
Total Transfer Credits	.024	.012	.060	1.918	.056
Age at the time of testing (years)	.018	.037	.015	.488	.626
Gender = male	3.986	.825	.142	4.834	< .001*
Major = Accounting	.969	.919	.031	1.054	.292
Major = Business Law	2.588	1.624	.047	1.593	.111
Major = Human Resources Management	-2.394	2.641	-.026	-.906	.365
Major = Marketing	.814	1.518	.016	.536	.592
Major = Other	6.627	2.046	.096	3.238	.001*
Ethnicity = African American	-3.519	1.535	-.156	-2.293	.022
Ethnicity = Hispanic/Latino	-7.235	2.171	-.129	-3.332	.001*
Ethnicity = Caucasian	-1.054	1.568	-.043	-.672	.502
Ethnicity = Other	-3.696	2.186	-.067	-1.691	.091
Normalized degree completion time(years)	.081	.084	.029	.970	.332

The final variable detailed in question 1, sub-question 1 as warranting additional investigation regarding interaction was ethnicity. As indicated in Table 4.1, the only ethnic categories with sufficient sample membership to support separate statistical analysis were African American (n = 480) and Caucasian (n = 208). Repeating the

regression reported in Table 4.5 for only the African American sample students resulted in the coefficients reported in Table 4.8 below. Adjusted R^2 dropped considerably from the original model to .175. The coefficients for the African American students were similar to those of the sample as a whole. Percent of study at a distance had a beta of 6.4 and was strongly significant. A one-point increase in major GPA was expected to add over 4 points to the MFT-B score. All else being equal, males are expected to have a 4-point higher score than females. Transfer credits were no longer significantly non-zero, but otherwise the coefficients looked quite similar to those estimated for the sample as a whole.

Table 4.8. Regression Coefficients, African American Students Only

Model	Beta	Standard Error	Standardized Beta	t-score	p
(Constant)	109.938	3.209		34.256	< .001
Percent of local credits earned online	6.375	1.333	.221	4.784	< .001
Transfer credits earned	.013	.016	.038	.835	.404
GE GPA	2.549	1.277	.150	1.996	.047
Major GPA	4.221	1.469	.221	2.873	.004
Age at the time of testing (years)	.044	.048	.042	.920	.358
Gender = male	3.979	1.169	.146	3.404	.001
Major = Accounting	1.623	1.150	.061	1.411	.159
Major = Business Law	-.014	2.219	.000	-.006	.995
Major = Human Resources Management	-3.106	2.798	-.048	-1.110	.268
Major = Marketing	2.313	2.257	.045	1.025	.306
Major = Other	5.768	2.585	.097	2.231	.026
Normalized degree completion time(years)	.111	.097	.050	1.140	.255

Re-estimating the regression for only the white students resulted in an adjusted R^2 of .296 and the coefficients reported in Table 4.9 below. While percent of study at a distance was still strongly significant, its coefficient was below 6 points, lower than any model reported above except the male-only subset. Note that African American students in the sample were disproportionately female (86% versus 81% overall) while white students were disproportionately male (25% versus 19% overall). Gender fell out as a significant factor among the white subset. In a result comparable to that of the male subset, general education GPA is no longer significant and the major GPA coefficient increased in magnitude considerably. And once again, multicollinearity was as likely cause of the change as substantive underlying differences between the subset and the original sample.

Table 4.9. Regression Coefficients, Caucasian students only

Model	B	Std. Error	Standardized Beta	t-score	P
(Constant)	82.229	7.907		10.400	< .001
Percent of local credits earned online	5.706	2.162	.179	2.639	.009
Transfer credits earned	.068	.029	.156	2.335	.021
GE GPA	3.536	3.166	.118	1.117	.265
Major GPA	12.306	3.823	.350	3.219	.001
Age at the time of testing (years)	2.562	1.605	.093	1.596	.112
Gender = male	-.005	.079	-.004	-.070	.945
Major = Accounting	.869	2.208	.023	.394	.694
Major = Business Law	4.923	3.324	.088	1.481	.140
Major = Human Resources Management	9.081	10.336	.050	.879	.381
Major = Marketing	.539	3.142	.010	.172	.864
Major = Other	6.383	4.457	.086	1.432	.154
Normalized degree completion time(years)	.036	.247	.009	.146	.884

Regressions in which an interaction term between each of the African American, white, and Hispanic/Latino ethnic categories in turn and the impact of online study were also fit. So too was a model with an interaction term between age and the percent of DE, to make it possible for age to influence the consequence of online study, perhaps with older students being more self-disciplined and self-directed and thus better able to benefit from study at a distance. Interaction terms were in none of those cases significant at the .05 level. In sum, there was little empirical evidence that gender, age or ethnicity impacted the influence of studying at a distance. With the exception of the male subset, the coefficient for the percentage of study completed at a distance was consistently strongly significant and positive, ranging in magnitude from 7.2 points for the entire sample to 5.7 for the subset of white students only. And although one could readily posit how it might have an effect, age did not appear to impact the consequence of online study here.

Question One Summary

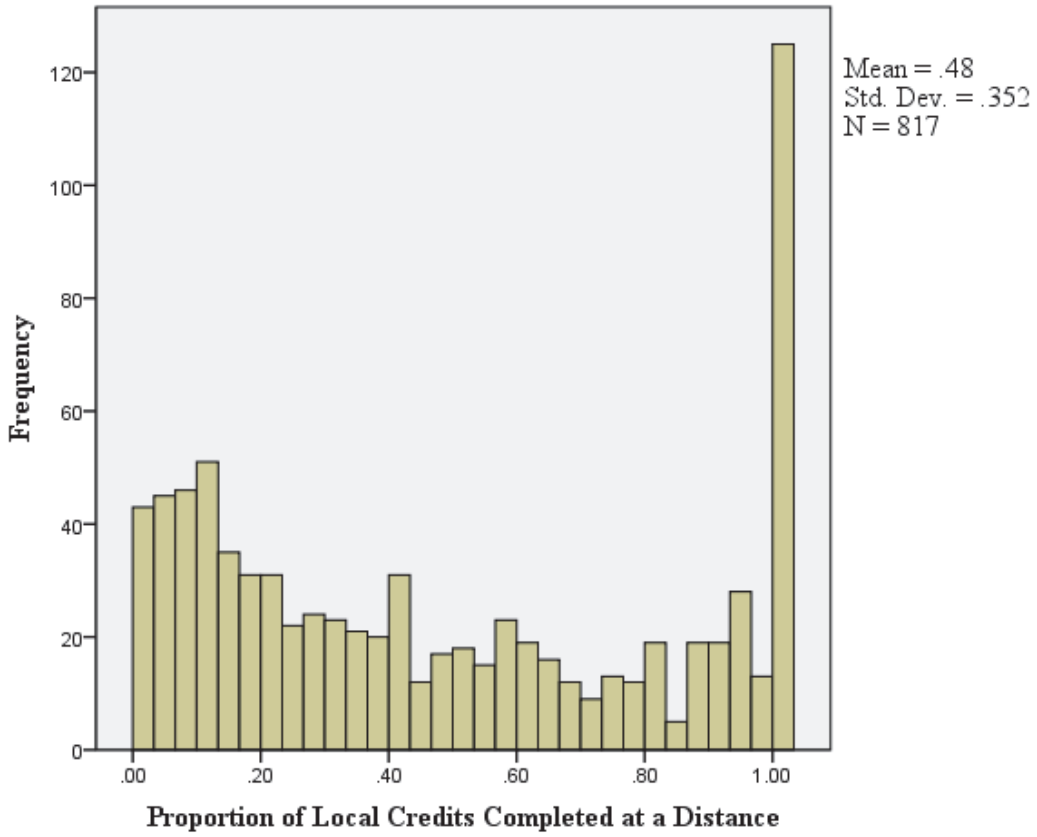
Completing one's studies at a distance appeared to have a strong, consistent, positive impact on expected MFT-B score. All else being equal, students studying entirely online were expected to have a roughly seven-point increase in their MFT-B scores than those studying entirely in the hybrid format where class-based meetings occur regularly. As a comparison of practical significance, those whose cumulative major GPA was a perfect 4.0 at the time of test taking were expected to do roughly 11 points better than those with a cumulative 2.0 major GPA (in most cases the minimum major GPA to qualify for graduation). From this comparison it is clear that studying at a distance was of significant practical consequence. As an indicator of the magnitude of a seven-point

change in MFT-B score in general, note that a seven-point change in mean institution-wide MFT-B score in 2011 was enough to move an institution's relative standing among all others a full 16 percentile points (Educational Testing Service, 2011a). Given that the MFT-B is an increasingly important indicator of undergraduate business program success (Martell, 2007), should this finding be generalizable beyond Urban College, it certainly would be of interest to deans and department chairs in business generally.

Results Excluding Students Studying Exclusively in One Format or Another

The full MFT sample had 817 subjects, 791 of whom had values available for all variables of interest. Of those 791, 26 studied exclusively in the classroom and 125 studied exclusively online. A histogram of the percentage of credits completed at a distance for all 817 MFT subjects is provided in Figure 1 below. The distribution is bi-modal, with a large contingent of students studying exclusively at a distance (a significant proportion of whom are sufficiently distant geographically for study on campus to be impracticable) and a second peak at roughly the 10% level representing a group of students for whom study at a distance is limited.

Figure 4.1. Histogram of the Distribution of Percentage of Local Credits Completed at a Distance



Because the influence of online study is of particular relevance for the subset of students who show an inclination to switch between formats, in the analysis that follows the regression reported in Table 4.3 above was re-estimated for just the subset of students who mixed both formats. First the equations were re-run exclusive of students studying entirely in one format or the other (N = 644, adjusted R^2 = .255, results are reported in Table 4.10 below). It was then re-run again, excluding students who took more than 80% of their credits in one format or the other (N=329, R^2 = .218, results reported in Table 4.11 below) and creating a subset of those who took a more even balance of both delivery formats.

Table 4.10. Regression Coefficients for Subjects with Percentage Completed Online less than 100 and greater than 0

(N=644)	Beta	Std. Error	Standardized Beta	p	VIF
(Constant)	108.853	3.655		< .001	
Percentage of total local credits that were earned online	6.583	1.347	.189	< .001	1.285
Transfer credits earned	.024	.014	.062	.089	1.157
GE GPA	2.463	1.211	.134	.042	3.737
Major GPA	5.439	1.406	.261	< .001	3.915
Age at the time of testing (years)	.044	.042	.040	.294	1.235
Gender = male	5.320	.933	.200	< .001	1.065
Major = Accounting	.924	1.026	.032	.368	1.075
Major = Business Law	2.082	2.049	.035	.310	1.027
Major = Human Resources Management	-2.272	2.652	-.030	.392	1.046
Major = Marketing	.508	1.593	.011	.750	1.117
Major = Other	1.785	2.772	.022	.520	1.049
Ethnicity = African American	-2.850	1.862	-.134	.126	6.601
Ethnicity = Hispanic/Latino	-5.814	2.557	-.109	.023	1.983
Ethnicity = Caucasian	-.437	1.934	-.018	.821	5.701
Ethnicity = Other	-2.611	2.475	-.053	.292	2.142
Normalized degree completion time(years)	.107	.096	.040	.268	1.117

Excluding the all-or-none students did not change the regression coefficients' signs, rough magnitudes or significance markedly from those reported for the entire sample in Table 4.3. The consequence of study at a distance appeared roughly equivalent between the entire sample and the subset that had at least experimented with both formats. So too did the impact of the other factors. The adjusted R^2 dropped from .311

to .255, although this was at least partly a result of the reduction in available observations. Thus there was little evidence that the influence of online study was limited to those students who opted to study entirely in one course format or the other.

Narrowing the sample further by limiting it to those whose credits were not more than 80% in one format or the other, however, resulted in entirely different coefficients and significance levels, as detailed in Table 4.11. Study at a distance here showed no significant influence, and its coefficient dropped from the 5-8 point range in previous models to .25 here. Indeed, outside of the constant, the only coefficient that remains significant at the .05 level was gender (p=.001). While this result might indicate that the impact of study at a distance was insignificant within more moderate bounds, it is quite likely that the marked reduction in available observations (N=791 versus 329) and its resultant reduction in statistical power was at issue, particularly given that several other, previously significant variables dropped off as well (for example the GPA, major and ethnicity variables). Nevertheless, it is certainly possible that for those who are sufficiently comfortable with both modes of study to switch readily between the two, the proportion of their study completed at a distance simply did not make a difference. One can by no means conclude that from these results, however.

Table 4.11. Regression Coefficients for Subjects Whose Percentage Completed Online was Between 20% and 80% Inclusive

	Beta	Std. Error	Standardized Beta	t-score	p	VIF
(Constant)	116.158	5.086		22.840	< .001	
Percentage of total local credits that were earned online	.246	2.939	.004	.084	.933	1.151
Transfer credits earned	.017	.019	.047	.923	.357	1.094
GE GPA	3.052	1.597	.179	1.911	.057	3.687
Major GPA	3.657	1.897	.188	1.928	.055	3.986
Age at the time of testing (years)	.043	.061	.039	.703	.483	1.305
Gender = male	4.474	1.345	.170	3.326	.001*	1.092
Major = Accounting	-1.019	1.273	-.041	-.800	.424	1.078
Major = Business Law	5.108	2.708	.095	1.886	.060	1.058
Major = Human Resources Management	-1.647	3.366	-.025	-.489	.625	1.053
Major = Marketing	-1.395	1.932	-.038	-.722	.471	1.169
Major = Other	-2.350	3.923	-.030	-.599	.550	1.028
Ethnicity = African American	-1.654	2.536	-.081	-.652	.515	6.444
Ethnicity = Hispanic/Latino	-3.756	3.431	-.076	-1.095	.274	1.994
Ethnicity = Caucasian	1.042	2.651	.044	.393	.694	5.252
Ethnicity = Other	-3.709	3.336	-.082	-1.112	.267	2.299
Normalized degree completion time(years)	-.047	.124	-.020	-.383	.702	1.156

Research Question 1, Sub-question 2

The second sub-question of research question 1 – "Are there individual courses whose grades are, all else being equal, more closely correlated with MFT-B scores than others?"

– involved investigating student grades at a finer grain that the general education cumulative GPA versus major cumulative GPA analyzed previously. Using course grades in a regression model similar to those discussed previously, however, presents several problems. First, it introduces the potential for major problems with collinearity. Student performance in courses whose subject matter is similar can be expected to perform similarly, introducing the strong possibility of high but stochastic pairwise correlation.

Data at the course level were also on the whole prohibitively sparse. Students in the sample took in total 515 distinct course numbers collectively, 130 of which were in business. Including even a modest subset of that total would have reduced the power of the resultant model dramatically. Further, while there is a core of business courses that all students are required to take, for most of the rest of the courses outside of that core, data available were quite disparate. That is, for many of these courses, only a very small subset of the student sample had actually taken them, creating large volumes of missing data.

For these reasons, modeling the influence of individual course grades was impracticable. Pairwise correlations between individual courses and MFT-B scores shed some light, and are reported in Table 4.12 below, but are suboptimal for the question at hand because each course is considered in isolation from all the others. Further, the influence of the other factors included in the regression equation is not accounted for statistically in the correlations. Nevertheless, pairwise correlations between grades in each of the core business courses and MFT-B scores were worth examining.

Table 4.12. Pairwise Correlations between final Course Grades in Core Business Classes and MFT-B Scores

Course	Correlation
Introduction to Business Systems	.201
Introductory Marketing	.177
Business Statistics II	.146
Business Ethics	.143
Introduction to Finance	.122
Organizational Dynamics	.115
Financial Analysis	.111
Business Statistics I	.107
Business Law	.095
Operations Management	.088
Marketing Management	.082
Accounting Principles I	.082
Human Resource Management	.066
Concepts in Leadership	.064
Accounting Principles II	.044
Macroeconomics	.044
Applied Management Concepts	.043
Microeconomics	.032
Policy & Strategy	.030
Concepts in Ebusiness	.020

All individual courses were positively correlated with MFT-B, although most only marginally. The basic introductory business course, MGT 101 – Introduction to Business Systems, is the most highly correlated at .201. If any course were to have been highly correlated, a broad, foundational business course is a likely candidate, although a more advanced course might also have made sense for different reasons. The next three courses most highly correlated courses include Introductory Marketing, the second Business Statistics course, and Business Ethics – a disparate group, subject-wise.

To gain a better sense of the collective impact of course grades on MFT without overwhelming the regression model with dozens of individual course grade variables, an exploratory factor analysis was conducted to identify GPA "factors" – clusters of related statistically related courses – that can then be included as a replacement for major and general education GPA back in the regression model discussed previously. Factor analysis reduces the dimensionality of the individual course grade data to a more manageable set of latent underlying factors. These factors are calculated to group highly correlated classes together in a given factor while at the same time minimizing the correlation across different factors. Factor analysis is based on the presumption that a large number of individual variables such as the individual course grades of the students in the sample set are based on a smaller set of latent, unobserved variables. For example, the circumstances that led one to do well in a given cost accounting class will likely also come into play for subsequent cost accounting classes, and to a lesser extent to all other accounting and finance classes.

Having had no *a priori* expectations regarding the number of resultant factors, I used the default SPSS extraction condition that the factor eigenvalues, which represent the extent to which variation in underlying variables is accounted for by the given factors, be greater than one. Missing values were deleted list-wise, that is, any record containing missing values for any one or more of the variables was eliminated from the analysis. A sample of 50 students who had taken all 20 core classes remained after list-wise deletion. Candidate course grade variables were limited to those business courses that half or more of the sample students had taken; there were 20 such courses. While this was not necessary for the factor analysis itself, it helped assure that the resultant factors were

calculable for a large portion of the students, which was important given that the factors were calculated so that they could be introduced back into the regression equation reported in Table 4.3, in lieu of the less specific "major GPA' variable. Extraction via principal components resulted in six factors, which together explained 77% of the variation in the underlying grades. Varimax rotation was employed to maximize the extent to which each grade variable loaded on one and only one factor, and because an orthogonal rotation method was sought to minimize correlation among factors. With a factor loading cutoff of .5, all 20 courses loaded on at least one factor and only four loaded significantly on two. The resultant factor matrix is presented in Table 4.13 below. Courses that loaded strongly (>.7) on a given factor are shaded in dark grey while those that loaded less strongly (>.5) are shaded in light grey.

Table 4.13. Rotated Course Grade Component Matrix

Course	Component					
	1. Quantitative	2. Leadership	3. Operations	4. Policy	5. Law & Ethics	6. Intro to Finance
Accounting Principles I	.855	.032	-.092	.032	-.154	.304
Accounting Principles II	.737	.517	.011	-.136	.113	-.002
Business Statistics I	.710	.188	.337	.252	.253	-.117
Business Statistics II	.689	-.082	.177	.503	-.070	.014
Introductory Marketing	.591	.388	.280	.031	.367	-.174
Financial Analysis	.546	.212	.324	.415	.021	.277
Introduction to Business Systems	.525	.096	-.576	.045	.354	.246
Concepts in Leadership	.129	.770	.099	.083	-.008	.325
Concepts in Ebusiness	.060	.746	.426	.154	.117	.039
Marketing Management	.203	.742	.058	-.010	.300	.117
Applied Management Concepts	.032	.164	.733	.048	.267	.102
Microeconomics	.266	.287	.601	.395	.013	-.264
Organizational Dynamics	.473	.320	.549	-.040	.072	.335
Operations Management	.305	.521	.524	.063	-.110	.044
Macroeconomics	.106	.042	-.002	.907	.227	.105
Policy & Strategy	-.037	.521	.214	.538	.128	-.483
Business Ethics	-.074	.080	.011	.165	.805	-.056
Business Law	.477	.258	.431	-.028	.538	-.021
Human Resource Management	.343	.152	.427	.247	.530	.361
Introduction to Finance	.119	.396	.059	.091	-.002	.831

Naming factors is always arbitrary, but the six resultant factors lent themselves fairly readily to heuristic classification. The sixth factor consisted of a single course,

Introduction to Finance, and thus required no further categorization. While using an eigenvalue of one as a threshold for factor elicitation does afford the possibility of a single-variable factor, it was not clear why this course did not load on the quantitative factor along with financial management and the accounting courses as opposed to loading by itself. Names selected for the factors identified can be found in the row headings in Table 4.13 above.

The six components identified through the factor analysis provided more specificity than the very broad major GPA employed previously while avoiding the dimensionality problems involved in individual courses, allowing a means to examine the influence of student grades in each of a number of performance areas and minimizing the correlation between those areas.

Average grades for the component courses in each of the six factors were calculated for each of the students in the sample and then these per-factor averages were used to replace major GPA in the original regression equation. The grade for each course was included only in the average of that factor upon which it loaded most highly irrespective of whether it loaded on additional factors as well. Re-estimating the regression equation with the six grade factors replacing major GPA resulted in an adjusted R^2 of .256 and coefficients reported in Table 4.14 below.

Table 4.14. Regression Coefficients with major GPA replaced with per-factor averages

	Beta	Std. Error	Standardized Beta	t-score	p
(Constant)	109.166	7.018		15.555	< .001
Percentage of total local credits that were earned online	6.157	1.971	.190	3.124	.002*
Transfer credits earned	.026	.025	.058	1.014	.312
GE GPA	3.415	1.838	.158	1.858	.064
Quantitative Factor GPA	6.018	2.081	.262	2.892	.004*
Leadership Factor GPA	2.034	1.406	.108	1.447	.149
Operations Factor GPA	-.957	1.830	-.042	-.523	.601
Policy Factor GPA	.389	1.245	.021	.312	.755
Law & Ethics Factor GPA	-3.627	1.347	-.199	-2.692	.008*
Intro to Finance Factor GPA	1.214	1.164	.068	1.043	.298
Age at the time of testing (years)	.049	.071	.040	.696	.487
Gender = male	4.440	1.525	.157	2.911	.004*
Major = Accounting	-.350	1.842	-.011	-.190	.850
Major = Business Law	8.696	3.164	.151	2.748	.006*
Major = Human Resources Management	2.039	4.617	.024	.442	.659
Major = Marketing	5.149	3.180	.089	1.619	.107
Major = Other	4.860	4.606	.058	1.055	.292
Ethnicity = African American	-4.527	3.248	-.199	-1.394	.165
Ethnicity = Hispanic/Latino	-10.260	4.542	-.170	-2.259	.025*
Ethnicity = Caucasian	-2.092	3.311	-.088	-.632	.528
Ethnicity = Other	-6.026	4.327	-.109	-1.393	.165
Normalized degree completion time(years)	.096	.321	.017	.298	.766

Variance Inflation Factor scores were under 3 for all coefficients except the African American and white ethnic categories, which were 7.4 and 7.1 respectively. Again, because these two mutually exclusive categories together represent over 87% of the sample total, pair-wise collinearity between the two is inevitable.

The GPA of the courses making up the quantitative factor was the best predictor of MFT-B scores by a large margin. This is consistent with results reported by Mirchandani et al. (2001) after a similar analysis of MFT-B correlates. Bagamery et al. (2005), however, found that their general business GPA factor predicted MFT-B better than their quantitative one.

Interestingly, the only GPA factor other than the quantitative one that was significant at $\alpha=.05$ was the Law & Ethics factor, which was in fact *negatively* associated with MFT-B score, meaning that increased performance in the courses comprising the law and ethics factor was associated with decreased performance on the MFT-B. A cynical sort might have expected an increased capacity for law and ethics to reduce success in the world of business generally, but it is nevertheless difficult to fathom why it would have led to lower performance on a test of business knowledge such as the MFT-B. Alternately, these courses may have required more qualitative skills, while the test focused on quantitative ones, and the negative correlation was an artifact of chance.

In sum, quantitative courses – particularly accounting and business statistics, together with foundational introductory courses in business and marketing, appeared more closely correlated with MFT-B scores than others, holding other factors constant. Overall business GPA, though, appears to be as good or better a predictor of MFT-B score than any of its component parts as identified through exploratory factor analysis.

Research Question 1, Sub-question 3

The third sub-question of the study's research question one pertained to the impact of DE-vs.-CI format in the change in grade from the first to the second of two-course sequences: in multi-course sequences (statistics 1 and statistic II, for example),

does a student's having taken the first course in the sequence in DE or CI format impact the grade expected in the follow-on? In such course sequences, the learning outcomes from the first of the sequence are presumed to be essential to success in the follow-on course. One would expect an individual receiving an 'A' in business statistics I to be better positioned to make an 'A' in business statistics II than someone who earned a 'C' in the first course. It is possible, though, that on average those who earned a given grade in the prerequisite class learned more (or less) than their counterparts in the classroom.

Focusing on the change between the grade in the second and the grade in the first course allowed a measure of relative achievement across the two. Because grade in the second course is presumed to be predicated on learning outcome mastery in the first, these sequences provided an opportunity to test if taking the first course at a distance impacts relative performance in the second. If having taken the first course in a traditional classroom format leads to significantly higher relative performance in the second course, one could argue that, holding absolute performance in the first constant, students were better prepared to achieve in the second, presumably because their learning in the first was in some way more profound or accessible in the subsequent course.

Given that accounting and business statistics appeared from the previous analysis to be among the subjects with the largest impact on expected MFT-B performance, and because they both are taken in required, two-course sequences at Urban, I selected these subjects for the grade differential analysis. The sequences examined were Accounting 101 and 102, Principles of Accounting I and II, respectively, and MAT 109 and 210, Business Statistics I and II. The sample was the subset of the 817 students having taken the MFT-B who formed the sample for the previous research questions and who also took

the course sequences in question. Five hundred twenty-eight of those students took both Accounting 101 and 102 at Urban College, while 445 took both Math 109 and 210.

In the first sequence analysis, the numeric grade in the first course was for each student subtracted from that of the second and recorded as the grade differential. Each student was coded as either having taken both classes at a distance, both in the classroom, first distance then classroom or first classroom then distance. For both the accounting and statistics sequences, an analysis of mean differences between those studying the first course in the sequence at a distance versus those studying in the classroom was conducted. Next those groups were further split into those who took the second course at a distance and those who took it in the classroom. Per sequence format counts are reported in Table 4.15 below.

Table 4.15. Format Sequence Distribution, Accounting and Statistics Grade Differentials

	Accounting		Statistics	
Format 1 to Format 2	N (percent)	Mean (Std. Dev.)	N (percent)	Mean (Std. Dev.)
Distance to Distance	87 (17%)	.22 (.57)	75 (17%)	-.11 (.78)
Distance to Classroom	6 (1.2%)	.25 (.92)	13 (3%)	.48 (.75)
Classroom to Classroom	405 (80%)	.07 (.72)	332 (75%)	-.02 (.71)
Classroom to Distance	10 (2.0%)	.23 (.65)	25 (6%)	-.40 (.85)
Total	508 (100%)	.10 (.70)	445 (100%)	-.04 (.74)

To test distance-first results against classroom-first results, an independent, two-sample t-test of mean differences was conducted. To evaluate all four sequence states (distance-distance, distance-classroom, classroom-classroom, and classroom-distance), a one-way, four-factor ANOVA was conducted. These results are reported per subject in the following.

Accounting 101 and 102

The average GPA difference between all students' accounting courses was .1 grade point, with a standard deviation of .70. For distance-first students, the average was .22 with a standard deviation of .59. For classroom-first students, the average difference was .07, with a standard deviation of .71. This difference when subjected to an independent sample t-test was not significant at the .05 level assuming equal variances ($p = .068$) but was if equal variances were not assumed ($p=.041$).

Four-group one-way ANOVA allowed for the categorization not only of the format of the initial class in the sequence but also the follow-on. The ANOVA resulted in non-significant between-groups difference at the .05 level ($F=1.29$, $p < .277$) and thus no significant differences between any of the four format sequences (classroom-classroom, classroom-online, online-online, or online-classroom).

Treating the grades in the two courses as repeated measures of achievement in accounting rather than examining the difference in grades between the two allowed for further analysis. The average grades for each of the two courses for the four format pairs are reported in Table 4.16 below.

Table 4.16. Descriptive Statistics for Accounting I & II Repeated Measures Analysis

	Format	Mean	Std. Deviation	N
Accounting 101	Both Classroom	3.3368	.75790	405
	Classroom to Distance	3.2800	.48944	10
	Both Distance	3.4080	.62547	87
	Distance to Classroom	2.9833	.73052	6
	Total	3.3437	.73209	508
Accounting 102	Both Classroom	3.4077	.65500	405
	Classroom to Distance	3.5100	.42544	10
	Both Distance	3.6264	.58179	87
	Distance to Classroom	3.2333	.75277	6
	Total	3.4451	.64456	508

Scores in Accounting 101 for those who took it at a distance but then switched to take 102 in the classroom were considerably lower than average (2.98 compared to 3.3), which may be driven by an inclination among those who did not do as well as they had hoped taking the first accounting class at a distance to take the second one in the classroom. Repeated measures ANOVA shows the effect of the within subjects factor (i.e. the difference between Accounting 101 and 102) to be significant at the .05 level ($p = .037$) but not the effect of that level together with the formats ($p = .277$). Similarly, Tukey's HSD post-hoc test shows no significant pair-wise differences between the four format pairs.

Limiting repeated measures comparisons to classroom-first versus online-first showed similar results. The within-subjects factor's impact – that is, the difference in grades from the first to the second accounting course – was significant at .05 ($p<.001$),

but the impact of that factor times the two formats was not (p=.068). Those who studied the first accounting course online did comparatively better in the second course compared to the first than those who studied the first at a distance did, but not to a statistically significant extent. While there was no evidence that taking the first accounting course at a distance prepares students better than taking it in the classroom, nor did it appear to hurt.

MATH 109 and 210

The average GPA difference for all students' statistics classes was -.04, with a standard deviation of .74. For distance-first students it was -.03 (.80 std. dev.) as compared to -.04 (.73 std. dev.) for classroom-first students. An independent samples t-test showed this difference to have no statistical significance at $\alpha = .05$ ($p = .848$ assuming equal variances, .856 if not). While the differences were not significant for the accounting classes either, the pattern is the same in both math and accounting, albeit with a thinner margin in math: the GPA change was larger for distance-first students than it was in classroom-first students, although, again, to a statistically negligible extent.

Breaking the two groups down further to reflect the format of the follow-on Math 210 as well as the initial Math 109 and then conducting a four-group ANOVA showed some significant differences between the more specific groups. The ANOVA's resultant between-groups differences were significant at the .05 level ($F=4.5$, $p=.004$). The Levene statistic did not provide evidence of heterogenic variances at $\alpha = .05$ ($p=.44$), so the post-hoc analysis was conducted via the equal variance-assuming Tukey's honestly significant differences (HSD) test. Using that test, the following mean differences were significant at the .05 level: between distance-to-classroom (+.48 mean change) and each of the

following other two format pairs: classroom-to-distance (-.40 mean change, p=.003) and distance-to-distance (-.11 mean change, p=.035). These results made the choice of format for MAT 210, the second course in the sequence, appear more consequential than the choice of format for the first, MAT 109. This could have been a result of a tendency for students who struggled with math at a distance in the first Math 109 course resolving to improve their odds by taking the follow-on Math 210 in the classroom, and doing better in the second course as a result, compared to the opposite result for the opposite switch. Such explanations are speculation, however. Broader investigation of these patterns in the future is warranted.

As with the accounting courses above, treating the grades in the two math courses as repeated measures of achievement in accounting rather than examining the difference in grades between the two allows both levels to be dependent variables impacted by format sequence. Descriptive statistics for both course levels are reported in Table 4.17 below.

Table 4.17. Descriptive Statistics for Math I & II Repeated Measures Analysis

	Format	Mean	Std. Deviation	N
MAT109	Both Classroom	3.25	.727	332
	Classroom to Distance	3.14	.718	25
	Both Distance	3.39	.696	75
	Distance to Classroom	2.82	.607	13
	Total	3.26	.723	445
MAT210	Both Classroom	3.24	.713	332
	Classroom to Distance	2.75	.841	25
	Both Distance	3.28	.717	75
	Distance to Classroom	3.31	.626	13
	Total	3.22	.726	445

Those students who started statistics at a distance and switched to the classroom had a markedly lower average GPA for the first course than average (2.82 versus 3.26). This fact may have motivated the switch from studying at a distance in the first to opting for the classroom in the second. If so, it appears to have paid off, given that the Math 210 average for those same students was higher than the overall average (3.31 versus 3.22 overall). A two-way repeated measures ANOVA was conducted to determine which of these results appeared statistically significant. That ANOVA indicated that the level of within-subjects dependent variable (here the first statistics course versus the second) was not a significant factor in the expected grade average ($F = .025$, $p=.874$), but that that factor together with course format sequencing was at $\alpha=.05$ ($F-4.53$, $p = .004$).

A post-hoc analysis was conducted to determine which course sequence pairings were significantly different from one another. Since the Levene test of equality of error variances was not significant at the .05 level for either factor level ($p=.732$ for Math 109 and .639 for Math 210), I used Tukey's HSD for the post hoc. The only resultant mean difference that was significant at the .05 level was that between those students who started in the classroom for their first statistics class and then switched to the distance format and those who took both classes at a distance ($p = .034$). Those who started in the classroom and switched to online did on average .39 GPA points worse than those students who completed both courses online. None of the other comparisons was significant. Repeating the repeated measures design but reducing the course sequencing classification to a binary distance-first versus classroom-first coding resulting in neither the within subjects factor level (Math 109 versus Math 210) nor the combination of that

level together with the binary course format being significant at the .05 level (p -.425 and .848 respectively).

In sum, there was little evidence that the decision to take the first course in a two-course sequence in accounting or statistics in one or the other format at Urban made any difference in itself in the expected performance on average in the subsequent course, with the exception that it appears that those who started in the classroom in Statistics I did worse switching to studying Statistics II at a distance that those who completed both courses at a distance. Those who went from distance to classroom improved, but they were also those whose initial performance most called for improvement. Revisiting the question of sequencing impact for individual subject sequences in the future for a larger sample or for a broader range of course sequences would be worthwhile, particularly for one involving a course sequence less technically rigorous than either math or accounting. Qualitative research on the motivations behind switching formats would be useful as well. Particularly in statistics, one sees the potential for the switch from statistics I at a distance to statistics II in the classroom to have been motivated by a desire for better instructor access and attention in statistics II in light of unsatisfactory grades in statistics I (2.82 for that group compared to 3.26 overall). At the same time, one could posit that the relatively higher performance in statistics II is the result of DE's being a better foundation at a given statistics I grade level for subsequent achievement in statistics II. These two scenarios are not especially compatible. Qualitative work may help explain what is motivating students who switch formats mid-sequence.

Research Question 2

The second main research question that was under investigation here pertains to the role studying at a distance has on the likelihood of retention to graduation. Specifically, it asks, "holding all other known relevant variables such as GPA, age, race, gender, major, and semester standing constant, does the extent to which students in a bachelor's program in business study at a distance as opposed to a traditional classroom impact their expected retention to graduation?" With regard to the first research question, there was evidence that the extent of online study was associated with increased performance on the MFT-B test, holding all other known factors constant. There was also some evidence in the literature that study at a distance was associated with reduced retention (Carr, 2000; Xu & Jaggars, 2011a, 2011b).

To the extent that, holding other relevant factors constant, an increased proportion of study at a distance is associated with a student's having a reduced likelihood of being retained to graduation, then this fact might to some degree explain that association. If, say, 'B'-average students mainly studying online are less likely to be retained to graduation than 'B' students studying primarily in the classroom, then perhaps those who do endure are disproportionately disposed to higher performance on standardized tests as well.[6]

To investigate that possibility among Urban College students similar to those in the original MFT-B sample, I developed a logistic regression model that estimated the likelihood of retention to graduation as a function of the percentage of local credits

[6] Note that actual retention differentials between the actual sample participants who favor DE and those who favor CI, a methodological concern in some DE-CI studies, is essentially a non-issue here because all students are in their final class before graduation.

completed at a distance, GPA, the number of credits transferred to Urban, gender, age, and ethnicity.

The original MFT-B sample was inappropriate for investigating questions of retention to graduation because students taking their MFT-B were in their final semester of study and thus all but certain to graduate. As such, a sample of students whose start dates afforded adequate time for graduation to date and yet was to the extent possible contemporary with the original MFT-B sample was sought. To strike this balance, students entering the business program in the years from 2000 to 2002 were used for the retention sample.

As with the MFT-B sample, interest here was primarily in student retention results for the four-year program. However, the nature of Urban College and its student information system precluded using those data. Urban business programs are "2+2", meaning that any student pursuing a bachelor's degree at some point along the way also confers an associate degree as well, irrespective of whether he or she has any specific interest in such a degree or not. Business students who had not yet met the graduation requirements for an associate degree were in the relevant period coded in the student information system as associate degree candidates whether or not they intended at the outset to pursue a bachelor's or not.[7] This precluded knowing which subset of those students who failed to persist to the associate degree level had also failed to persist to the bachelor's level versus those who had no intention to pursue a bachelor's. To the extent that those who fail to persist to a four year degree do so mainly in the first half of their studies anyway, the issue is moot, but in either case, the only retention to graduation

[7] This situation has recently been remedied and students' ultimate intentions are coded upon their entry. Those data are too recent, however, to afford a retention analysis.

analysis that the student information system afforded was that to the associate level, and thus that was the basis of the analysis that follows.

Demographic information and graduation status for all students who started studying in the Urban business program between the years of 2000 and 2002 were collected, together with transcript information to determine GPA, credits transferred from other institutions and proportion of studies completed at a distance. For graduation status, students were coded as a '1' if they had graduated at the time of retrieval (March 2012) and '0' if they had not (allowing an arbitrary but extremely generous 10-12 year span in which to graduate with an associate degree). A logistic regression was then fitted to these binary graduation status results, using GPA, percentage of local credits taken online, number of transfer credits, and demographic information as predictive variables. The resultant equation estimated the likelihood of graduation given the values of the predictive variables.

Of the retention sample's 768 subjects, 387, or 50.7% graduated by the time the data were extracted. Data were available for all variables for all subjects, except for gender, which was unavailable for five subjects, who were as a result excluded from estimation as a result. Descriptive data for the sample are summarized in Table 4.18.

Table 4.18. Descriptive Statistics for the Variables used in the Retention Equation

Variable	Mean	Std. Deviation
GPA	2.9	1.03
Transfer Credits	6.7	10.39
Graduated	51%	N/A
Gender = male	26%	N/A
Age	30.3	9.77
Percentage of credits completed at a distance	39%	.40
Ethnicity:		
African American	51%	N/A
Hispanic/Latino	5%	N/A
Caucasian	30%	N/A
Unknown	9%	N/A
Other (Non-citizen, Native American, or Asian American)	5%	N/A

Model coefficients and odds ratios for the retention equation are reported in Table 4.19 below. Betas here represents the natural logarithm of the marginal impact of a given variable on the odds of a positive outcome of the dependent variable (here retention to graduation). Because of the difficulty of interpreting beta, the odds ratio, which is calculated as the base of the natural logarithm (i.e., *e*) raised to the beta power, is provided as well. The odds ratio is the change in the odds of a positive outcome given a marginal change in the variable in question. For example, a one point increase in GPA increases the odds of retention to graduation 5.6 fold. An odds ratio of less than one (equivalent to a negative beta) reduces the odds of a positive outcome.

Table 4.19. Coefficients and Odds Ratios for the Retention Equation

	B	Odds Ratio (e^{β})	p
Percent of credits completed at a distance	-.846	.429	.001*
Cumulative GPA	1.724	5.609	< .001*
Transfer Credits	.022	1.022	.013*
Gender = Male	-.348	.706	.088
Age at the start of study (years)	-.024	.976	.018*
Ethnicity = African American	1.710	5.528	< .001*
Ethnicity = Hispanic/Latino	1.306	3.690	.012*
Ethnicity = Caucasian	1.432	4.189	< .001*
Ethnicity = Other	1.154	3.172	.036*
Constant	-5.678	.003	< .001*

To illustrate the influence of studying at a distance on expected retention here, taking the mean results for all the other predictor variables (or modal value for the categorical ones, to present to the extent possible a "typical" Urban business student), the model predicted that a student who studied entirely in the classroom has a 61% chance of graduating while one who spilt classroom and distance equally has a 50% chance, compared to someone studying entirely online, who has a 40% chance. As such, one would expect an 'average' Urban student from the sample who studied entirely in the classroom to persist to graduation (given that the odds at 61% are considerably better than 50-50 cutoff), while for the same student studying entirely online, one would not expect retention to graduation. From this perspective, proportion of local study conducted at a distance certainly has practical significance.

Other coefficients with positive impact significant at the .05 level included GPA (the largest influencer of graduation likelihood by far), amount of credit transferred to Urban, and each of the four ethnic categories (African American, Hispanic/Latino, Caucasian and other) relative to the referent category, "unknown". Interestingly, all else

being equal, African American students are more likely to be retained to graduation than white students, who in turn are more likely than Hispanic/Latino students. Urban's being a predominantly African American institution may play a role there.

Being male reduced the likelihood of graduating, but not to a statistically significant extent. Age at the start of study had a negative impact on retention likelihood that was statistically significant at the .05 level, but whose practical import was relatively moderate: All else being equal, being a decade older at the start of study reduced the odds of graduating by only slightly over 20%.

The classification model performed relatively well on the data from which it was originally fit: overall, 75.2% of observations were correctly classified by the model, including 82.9% of those who graduated and 67.3% of those who did not. As a mechanism for comparing the goodness of fit of the classification model to that of the MFT-B models discussed previously, the Cox and Snell pseudo-R^2 (a relatively conservative measure in that its maximum value cannot reach 1) for the logistic model is .311.

Summary

Studying at a distance was found to be positively correlated with MFT-B score while holding all known additional correlates constant. The positive impact of distance education demonstrated both statistical and practical significance. The influence was seen to a roughly equal extent whether the extent of online study was expressed as a credit total or as a percentage of credits completed locally. (The two are not equivalent because the number of credits transferred to Urban – itself a predictive variable – varies

widely among the sample students). There was little empirical evidence that gender, age or ethnicity impact the influence of studying at a distance. While by no means conclusive, there was some evidence that the impact on MFT-B score of studying at a distance may be less marked for male than for that of female students.

Conducting a factor analysis to identify groups of individual courses whose grades tended to be highly correlated relative to the others and then using the subject GPA averages for those factors as inputs to the MFT-B predicting regression model showed that quantitative courses like accounting and business statistics collectively had the biggest impact on MFT-B scores. An analysis of the instances of multi-course distance sequences involving those subjects followed, seeking evidence that the /classroom format of the first course in the sequence has an impact on the expected change in grade from the second course to the first. No evidence of such an impact was found.

Finally, the impact of study at a distance on retention to graduation was examined by fitting a logistic regression that predicted likelihood of graduation based on credits obtained at a distance, GPA, credits transferred into Urban and demographic factors. Here study at a distance was found to reduce the likelihood of retention to graduation significantly, holding the other variables constant.

In sum, study at a distance appeared to lead to a "survival of the fittest" condition whereby students studying at a distance were winnowed out more readily than their classroom-based counterparts. However, those who persisted – perhaps in part because of their having survived DE's more intense retention challenges –performed markedly better on the MFT-B test.

CHAPTER 5
CONCLUSIONS AND RECOMMENDATIONS FOR FUTURE RESEARCH

This concluding chapter summarizes the results reported in Chapter 4, discusses the implications of those findings and their relationship to the extant literature, discusses the study's limitations and recommends areas for further study.

Summary of Results

The first major finding of this study was that, holding other relevant factors constant, the extent to which students completed their studies at a distance was associated with improved student learning outcomes as measured by performance on the MFT-B exam, to a statistically and practically significant degree. All else being equal, those Urban students who completed more of their studies at a distance performed better on the MFT-B compared to their counterparts favoring traditional classroom-based study.

GPA, particularly for business-related courses but also for the general education core was strongly associated with increased MFT-B performance. This result is unsurprising ex-ante and in keeping with the empirical literature. Regarding more specific, course-level predictors of MFT-B performance, quantitative and core introductory business courses were in general the best predictors of MFT score. Holding all other factors constant, male students outscored women. Evidence of interaction between the impact of study at a distance and demographic variables such as gender, age or ethnicity was limited.

While the extent of study at a distance was positively associated with MFT-B score, it also predicted reduced likelihood of retention to graduation. Those studying

exclusively at a distance were less than half as likely to have graduated as those studying exclusively in the classroom.

Activity Theory and the Study Results

To the extent (potentially limited, as taken up in the section that follows) that higher expected MFT-B scores can be taken to imply improved learning generally, then this study may be seen as providing evidence that distance education is more effective (holding aside for the moment the fact that studying at a distance also appears to negatively impact retention) than hybrid online-plus-classroom education. While the notion that the entirely asynchronous online format is actually less effective when supplemented with regular class meetings is counterintuitive, through the lens of Activity Theory (AT), this result is plausible.[8] In AT, learning involves the student's broadened and increasingly productive intellectual and social involvement as mediated by technology and tools. While social involvement would on its face seem limited at a distance relative to on campus, involvement may tend to be more profound, and engagement with technological systems more pervasive as well. The "sage on the stage" mindset and the passivity that it can engender in students are avoided altogether in asynchronous DE. Successful distance students may feel compelled to take ownership of their education in ways and to extents different than students who favor the classroom.

[8] Note, though, that as a meta-theory – that is, a framework for developing theory, rather than a specific theory of learning itself – Activity Theory could also form the basis for expecting the opposite result as well, probably just as readily. Indeed, the nature of the social interactions afforded at a distance and the resultant norms, values and roles might be inadequate to support the caliber of learning that happens in the classroom. This is often either the chief appeal or shortcoming of a meta-theory, depending on one's intents.

That fact may cut both ways, though, and while successful study at a distance may be associated with deeper engagement or personal ownership, it may also *require* deeper engagement or ownership (as suggested by, for example, Bernard et al. (2009)). In either case, this study found that holding other known factors constant, students studying mainly at a distance are less likely to graduate than their campus-based counterparts.

Whether this is a result of distance education's increased demand for personal commitment and drive to engage one's studies or something else entirely, the result may help explain increased MFT-B scores for students favoring distance education. Modeling GPA and demographic characteristics by themselves (i.e. in the absence of course format history) may – as the data in this study indicated – understate the MFT-B scoring potential of the subset of students who exhibit the qualities necessary to overcome whatever persistence challenges studying mainly at a distance provides. The increased capacity to make it through when studying at a distance may also contribute an improved capacity for MFT-B performance, all else being equal.

Relationship of the Study Results to the Literature

Situating this study's findings within the context of the relevant literature is challenging for two reasons. First, literally tens of thousands of empirical DE-CI comparison studies have been published, with results ranging from dramatically better performance in the classroom to dramatically better performance at a distance. Meta-analyses of these studies, while hampered by the original studies' methodological inadequacies and problematic results reporting, are also common in the literature. These generally report a similar, albeit slightly narrowed range of results, from no significant

DE-CI difference (Bernard, Abrami, Lou, Borokhovski, et al., 2004; Cook et al., 2008), through to improved results at a distance (Allen et al., 2004; Jahng et al., 2007; Sitzmann et al., 2006).

A defining meta-analysis by the U.S. Department of Education Office of Planning Evaluation and Policy Development (2010) found that DE students performed "modestly" better than those in the classroom, but also added a critical refinement: students in programs that blended face-to-face interaction and online study had a greater outcomes advantage over pure classroom instruction than did pure distance-based classes. Students in purely online courses, on the other hand, performed equivalently to those studying strictly face-to-face. By extension, classroom plus online appeared to outperform online only study, at least relative to pure classroom work. By a sort of transitive property, the U.S. Department of Education study arrives at the conditions studied here directly: classroom-based study supplemented with the tools traditionally associated with online study versus those tools alone, devoid of face-to-face interaction. The Department of Education study indicated better performance in hybrid courses than in DE alone. The results here, however, are contrary to the Department of Education study, and are quite surprising: adding regular face-to-face interaction to the tools, technologies and interactions associated with modern, web-based distance education appears to reduce learning outcome achievement as measured by MFT-B scores.

In terms of CI-DE retention differences, results are again wide-ranging but the results generally report a negative impact of DE on retention. For example, Moody (2004) investigated the potential reasons why DE attrition is so high. Similarly, Bernard, Abrami, Lou, and Borokhovski (2004, p. 404) in their meta-analysis report a statistically

significant negative effect size for asynchronous DE on retention-related outcomes and also cite a "traditionally high dropout rate in DE". They also found no significant trend in the retention impact size over the time period they studied (1985-2002), suggesting no evidence that students were acclimating to whatever retention challenges DE presented. Xu and Jaggars (2011a) also report lower retention rates in DE compared to CI.

In her DE-CI meta-analysis of allied health programs, Williams (2006) reports that studies whose students are older report effect sizes that more significantly favor DE than those studies that involve younger students. While there was no evidence within the Urban student sample that age impacted MFT-B score holding the other factors constant, it is possible that the fact that urban caters to working adults and has in general an older, "nontraditional" student population may be a factor in DE students' outperforming those taking hybrid classes. In any case, it is somewhat surprising and contrary to the few comparable investigations reported in the literature.

Limitations

Like much of the DE-CI comparative literature, this study drew its data from a single institution. As such, generalization to other institutions or higher education generally is problematic. The marked, consistent evidence of strong heterogeneity of effect size in virtually every DE-CI comparative meta-analysis cited in this study illustrates this fact clearly. Either the studies making up these meta-analyses are not measuring the same construct in equivalent ways, or the studies' respective local conditions result in qualitatively different outcomes, or both. More widespread use of published instruments such as the MFT-B used here would certainly aid inter-study

comparability. However, objective, multiple choice tests like the MFT are – as detailed in the "limitations" section below – susceptible to criticism in terms of the extent to which they capture the sort of learning that matters. Also as discussed below, published, broadly normed instruments need not imply multiple choice questions of rote learning.

Additionally, general cognitive/test-taking aptitude is not as well modeled here via general education core GPA as in other research where SAT (e.g. Allen & Bycio, 1997; Bagamery et al., 2005) or GRE scores (Frazier & Edmonds, 2002) or ACT scores (Contreras et al., 2011) perform that function. Given the extent to which general aptitude factors tend to be central in the models that include them in predicting MFT performance, this limitation does give pause.

Two factors mitigate that limitation, to at least some degree. First is the fact that general education core GPA is in most of the models considered in this study a useful explanatory variable in its own right, albeit one whose influence is difficult to disaggregate from that of other GPA components like major GPA, which presumably capture subject matter mastery more so than general intellectual or academic wherewithal. The challenge of separating these respective impacts is by no means unique to the study's use of multiple GPA measures, and is as much an issue when modeling SAT, GRE or ACT together with GPA.

The second factor mitigating against the absence of a variable for general test-taking aptitude is that study at a distance has been here measured as a matter of degree, and the students taking their courses at a distance are not qualitatively different from those in the classroom; indeed, in many cases they are one and the same. This does not, however, preclude the possibility that general aptitude and the inclination to study at a

distance are not somehow related. The fact that excluding those studying exclusively in one format or the other did not change the model significantly would seem to indicate that such a relationship is not at issue, but the fact that the same did not hold when excluding those studying 80% or more in one format or the other does not.

Finally, while Urban provided an environment with considerable consistency between DE and CI sections, no effort was made to assess or control for variations in the quality of pedagogy between the two. Syllabi, texts, outcomes, course durations, and assessments were consistent and the pool of instructors was the same for the two, but pedagogy in one format might have been in a systematic way closer to the state of the art than in the other, and no effort was made to control for such deviations.

Future Research Recommendations

The recommendations for future research resulting from this study include the following:

1. Better Isolation of the Value Added by Student Learning

Controlling for general aptitude is paramount to separating out those gains in student learning outcomes that are attributable to what the students learn as a result of their studies from those that are a function of innate ability or preexisting knowledge. One approach to achieving this isolation is random assignment. A true experiment's random assignment to control and treatment groups can obviate this concern, but it simultaneously injects a measure of irrelevance: how well someone artificially induced to study in one format or another does there is of much less interest than the differences among those choosing freely. But in the absence of random assignment, one needs to

account for the influence of general test-taking aptitude somehow, and modeling it explicitly makes it difficult to disaggregate its influence from that of variables that capture learning, like GPA measures or content-knowledge based standardized test scores.

An alternative that offers advantages to random assignment is to make the value of students' studies more explicit by measuring learning longitudinally. The MFT-B test also exists in a version intended to be administered at the end of an associate degree program in business. Urban College is a "two plus two" institution and as such also awards associate degrees in business. Indeed, starting in 2010 all business majors, including those enrolled in the bachelor's program have been required to sit for the associate MFT at the end of their sophomore year. While such data are not yet available for more than a handful of students, when they are, it will afford a longitudinal, "value-added" perspective on MFT-B score over the latter half of the program that will make issues related to accounting for general test-taking aptitude moot. By including the associate MFT score among the independent variables predicting MFT-B, test-taking aptitude should be a wash, and whatever variation that can be attributed to the other predictor variables should be as a result of learning over the last two years of the four year business program.

2. More Studies that Focus on the Classroom's Contribution

DE-CI comparative studies have historically compared students studying in a traditional classroom context devoid of technology to technology-based distance education devoid of classroom interaction. This comparison has made sense in the past but is less reflective of current practice. Increasingly colleges and universities are

making the tools and systems rooted in distance education available for all courses, whether they have regularly scheduled classroom interaction or not. Over time one can expect most every classroom-based course to become what is currently referred to as a hybrid course. In keeping with the "commodification" of IT predicted by Carr (2004), a common set of core technologies currently provided by learning management systems such as Blackboard will likely be made available routinely in all classes. Differences in feature sets from one learning management system to the next will diminish over time. At the same time, online courses conducted asynchronously and entirely at a distance will likely continue to grow. Thus, the appropriate question becomes not what impact the presence or absence of technology has, but rather what the impact the presence or absence of *classroom interaction* has.

The meta-analysis produced by the U.S. Department of Education Office of Planning Evaluation and Policy Development (2010) afforded this comparison indirectly, by comparing the relative difference of strictly asynchronous online study versus the classroom and hybrid courses versus the classroom and found the achievement differential to be more marked for hybrid courses. This study made the hybrid versus distance comparison directly and saw results contrary to those of the Department of Education study. The Department of Education outcome is on its face more consistent with intuition: taking the set of tools and interactions that make up an asynchronous online course and supplementing them with same-time, same-place interaction improves learning outcomes. The contrary result observed here might be explained by the impact of regular class meetings on a student's sense of agency or "ownership" of his or her own learning. The positive impact of student control and direction on learning has been

reported at least as far back as Newman (1957). Irrespective of any potential causes for the study's results, further research on the impact of classroom interaction, with technological mediation constant, is surely warranted. Clearer delineation of the exact nature of hybrid classes than historically seen in the DE-CI-Hybrid comparative literature would be useful as well.

More consistently "apples-to-apples" comparisons are warranted too. Urban here presented a very convenient lab for such comparisons given that the same pool of instructors teaching the same content using the same syllabi, outcomes, assessments and technologies to teach both hybrid and exclusively asynchronous sections of the same classes to a pool of students who self-select between the two formats on a course-by-course basis, but who also switch readily back and forth between the two. Such a circumstance may be rare currently, but likely will grow increasingly common in the future.

3. More Authentic Outcomes Assessment

The MFT-B is an objective, multiple choice examination and as such is susceptible to the criticism that it is not an authentic measure of student achievement. Mason et al. (2010), for example, decry the MFT-B test as revealing little useful information about student performance, particularly relative to the cost of administering the examination. Mirchandani et al. (2001, p. 52) suggest that performance on standardized tests like the MFT-B do not correlate well with "actual knowledge or the ability to perform." There is evidence that the nature of the learning one seeks to measure does indeed impact the outcome when comparing DE to CI. For example,

Sitzmann et al. (2006) found DE comparatively better versus CI for declarative knowledge than for procedural knowledge.

At the same time, more authentic methods of assessment typically involve considerable tradeoffs. First, locally developed, subjective assessments do not normally afford the inter-institutional comparisons that accountability-driven assessment demands. It is difficult to imagine the role of accountability-driven assessment diminishing in the foreseeable future. Secondly, while standardized assessment can be quite expensive, its alternatives are more expensive still, as even Mason et al. (2010) in their criticism of standardized assessment and the MFT-B particularly themselves concede. Developing, validating, and administering institution-specific instruments, particularly those requiring – as authentic assessment endeavors are wont to do – human scoring is much more expensive in terms of both time and money than standardized assessment.

Assessment via computer-based simulations may help bridge this divide. Simulations have the potential to provide assessment of more meaningful, nuanced tasks under more realistic circumstances. At the same time, they can be standardized for cross-institutional comparisons and broader validation and also afford automatic scoring and thus reduced administration costs. Bennett, Persky, Weiss, and Jenkins (2010), Gobert and Koedinger (2011), Iseli, Koenig, Lee, and Wainess (2010), Lawrence, Reed, and Locander (2011), and Nelson, Erlandson, and Denham (2011) all investigate computer gaming and simulation from an assessment and accreditation perspective. Conducting the sort of investigation documented here, but replacing MFT-B scores with an automatically scored business simulation would be a valuable potential area for future research.

More broadly, qualitative research about the actual nature of distance education should continue unabated, particularly given how dramatically that experience has changed over time and how student attitude toward and facility with technology have changed as well. Further efforts to understand the experiences of those who undertake study at a distance and fail to persist to graduation are indicated as well.

Conclusions and Implications for Policy and Practice

Media are not mere channels through which educational payload travels. Both common sense and modern empirical experience revolt at Clark's (1983) likening of the influence of instructional media on learning to that of the truck that delivers our groceries. Likewise, tools – technological and otherwise – are not values-neutral. As Benkler (2006) explains:

> Different technologies make different kinds of human action and interaction easier or harder to perform. All other things being equal, things that are easier to do are more likely to be done, and things that are harder to do are less likely to be done. (p. 17)

All other things are never really equal, however, and while a given technology has values and consequences – what Wellman, Quan-Haase, Boase, and Chen (2003) characterize as affordances and Winner (1986) more strongly as politics – its form does not predestine its function. Its function is emergent and sometimes impossible to predict, but it can be shaped, by policy as well as by changing mores and attitudes.

Higher education both occurs via and is situated in complex sociotechnical systems. These systems are likely to be more pervasively and deeply mediated by technology over time. Even if that pervasion is inevitable, the ways in which technology

is knit into higher education's fabric, what it supplants, and its consequences for learning are not.

Technology must not be studied in isolation: changes in technology change social valences, mores, expectations, and how we frame our respective roles in the systems we participate in. These changes in turn alter how and where technology is used (what Orlikowski (1992) calls the "duality of technology"), creating a complex system of influences that are difficult to disentangle.

To assess technology's impact– or that of any intervention for that matter – on education, one must be prepared to measure learning, as daunting or distasteful as that task may sometimes seem, and to measure it in ways that reflect the sort of learning that matters. If the learning that we value most is the process of socialization and acculturation, then perhaps no objective test will fill the bill. Perhaps the learning that matters for meaningful professional business practice is unreflected or badly distorted by MFT scores.

At the same time, resources are always constrained, and those resources reserved for assessment are unavailable to actual teaching and learning. Assessment needs to be affordable and efficient. Published instruments, especially those that are automatically scored, provide valuable economies of scale to assessment efforts. Even beyond their cost efficiency, such instruments have a desirable parsimony. A bottom-line numeric score is not only useful for broad accountability purposes, it is also the sort of thing rational agents in this age of information overload seek out. We should look to technology to aid us in developing new generations of assessments that maintain that economy while at the same time better capture the sort of learning we seek to capture.

Distance education can broaden access to higher education dramatically. It enables those for whom regular trips to campus would be an outright hardship – those geographically remote or with complicated family or job obligations – to attain their degrees. Further, it can reduce the cost of providing a college education. Classroom space and associated physical plant costs are significant. In an era of growing public unease about the cost of college and increasingly burdensome student debt levels, opportunities for controlling costs are critical. But such benefits still come with risk. We must continue to be vigilant in our efforts to assure students both online and in the classroom are achieving the outcomes we hold out for them to achieve and that our efforts to improve teaching and learning in both formats are unwavering.

REFERENCES CITED

Adams, J., & DeFleur, M. H. (2006). The acceptability of online degrees earned as a credential for obtaining employment. [Electronic Article]. *Communication Education, 55*(1), 32-45.

Allen, I. E., & Seaman, J. (2010). Learning on demand: Online education in the united states. Retrieved March 8, 2011, from http://sloanconsortium.org/publications/survey/pdf/learningondemand.pdf

Allen, I. E., Seaman, J., & Garrett, R. (2007). Blending in: The extent and promise of blended education in the united states. Retrieved September 5, 2011, from http://sloanconsortium.org/publications/survey/pdf/Blending_In.pdf

Allen, J. S., & Bycio, P. (1997). An evaluation of the educational testing service major field achievement test in business. [Electronic Article]. *Journal of Accounting Education, 15*(4), 503-514.

Allen, M., Bourhis, J., Burrell, N., & Mabry, E. (2002). Comparing student satisfaction with distance education to traditional classrooms in higher education: A meta-analysis. [Electronic Article]. *American Journal of Distance Education, 16*(2), 83 - 97.

Allen, M., Mabry, E., Mattrey, M., Bourhis, J., Titsworth, S., & Burrell, N. (2004). Evaluating the effectiveness of distance learning: A comparison using meta-analysis. *Journal of Communication, 54*(3), 402-420.

Allen, M. J. (2003). *Assessing academic programs in higher education*: Jossey-Bass.

Bagamery, B. D., Lasik, J. J., & Nixon, D. R. (2005). Determinants of success on the ETS business major field exam for students in an undergraduate multisite regional university business program. [Electronic Article]. *Journal of Education for Business, 81*(1), 55-63.

Banta, T. W. (2007). Can assessment for accountability complement assessment for improvement?(analysis). *Peer Review, 9*(2), 9(4).

Barneveld, A. v., & Strobel, J. (2009). When is PBL more effective? A meta-synthesis of meta-analyses comparing PBL to conventional classrooms. *Interdisciplinary Journal of Problem-based Learning, 3*(1). Retrieved March 16, 2011 from http://docs.lib.purdue.edu/ijpbl/vol3/iss1/4

Bean, J. P., & Metzner, B. S. (1985). A conceptual model of nontraditional undergraduate student attrition. [Electronic Article]. *Review of Educational Research, 55*(4), 485-540.

Benkler, Y. (2006). *The wealth of networks : How social production transforms markets and freedom*. New Haven: Yale University Press.

Bennett, R. E., Persky, H., Weiss, A., & Jenkins, F. (2010). Measuring problem solving with technology: A demonstration study for naep. [Electronic Article]. *Journal of Technology, Learning, and Assessment, 8*(8).

Bernard, R. M., Abrami, P., Lou, Y., & Borokhovski, E. (2004). A methodological morass? How we can improve quantitative research in distance education. [Electronic Article]. *Distance Education, 25*(2), 175 - 198.

Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., et al. (2009). A meta-analysis of three types of interaction treatments in distance education. [Electronic Article]. *Review of Educational Research, 79*(3), 1243-1289.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovsk, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. [Electronic Article]. *Review of Educational Research, 74*(3), 379-439.

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379-439.

Bethel, E. C., & Bernard, R. M. (2010). Developments and trends in synthesizing diverse forms of evidence: Beyond comparisons between distance education and classroom instruction. [Electronic Article]. *Distance Education, 31*(3), 231-256.

Bhattacharya, M. (2008). *Why technology innovations are still a cottage industry in education?* Paper presented at the Proceedings of the 2008 Eighth IEEE International Conference on Advanced Learning Technologies - Volume 00.

Black, H. T., & Duhon, D. L. (2003). Evaluating and improving student achievement in business programs: The effective use of standardized assessment tests. [Electronic Article]. *Journal of Education for Business, 79*(2), 90-98.

Blin, F., & Munro, M. (2008). Why hasn't technology disrupted academics' teaching practices? Understanding resistance to change through the lens of activity theory. [Electronic Article]. *Computers & Education, 50*(2), 475-490.

Bond, L. (2007). The case for common examinations. Carnegie perspectives. Standford, CA: Carnegie Foundation for the Advancement of Teaching.

Brown, G., & Wack, M. (1999, January 20, 2012). The difference frenzy and matching buckshot with buckshot. *The Technology Source* Retrieved January 20, 2012, from http://technologysource.org/?view=article&id=320

Brown, J. S., & Duguid, P. (2000). *The social life of information*. Boston: Harvard Business School Press.

Bush, H. F., Duncan, F. H., Sexton, E. A., & West, C. T. (2008). Using the major field test-business as an assessment tool and impetus for program improvement: Fifteen years of experience at virginia military institute. [Electronic Article]. *Journal of College Teaching & Learning, 5*(2), 75-88.

Bycio, P., & Allen, J. S. (2007). Factors associated with performance on the Educational Testing Service (ETS) major field achievement test in business (MFAT-B). [Electronic Article]. *Journal of Education for Business, 82*(4), 196-201.

Carliner, S. (2004). *An overview of online learning* (2nd ed.). Amherst, Mass.: HRD Press.

Carr, N. G. (2004). *Does it matter? Information technology and the corrosion of competitive advantage*: Harvard Business Review Press.

Carr, S. (2000). As distance education comes of age, the challenge is keeping the students. *The Chronicle of Higher Education, 46,* A39 - A41.

Cavanaugh, C. S. (2001). The effectiveness of interactive distance education technologies in K-12 learning: A meta-analysis. *International Journal of Educational Telecommunications, 7*(1), 73-88.

Cavanaugh, C. S., Gillan, K. J., Kromrey, J., Hess, M., & Blomeyer, R. (2004). *The effects of distance education on K-12 student outcomes: A meta-analysis*: Learning Point Associates/North Central Regional Educational Laboratory.

Chyung, S. Y., & Vachon, M. (2005). An investigation of the profiles of satisfying and dissatisfying factors in e-learning. [Electronic Article]. *Performance Improvement Quarterly, 18*(2), 97-113.

Clark, R. E. (1983). Reconsidering research on learning from media. [Electronic Article]. *Review of Educational Research, 53*(4), 445-459.

Clark, R. E. (1994). Media will never influence learning. [Electronic Article]. *Educational Technology Research and Development, 42*(2), 21-29.

Clark, R. E. (2000). Evaluating distance education: Strategies and cautions. [Electronic Article]. *Quarterly Review of Distance Education, 1*(1), 3-16.

Contreras, S., Badua, F., Chen, J. S., & Adrian, M. (2011). Documenting and explaining major field test results among undergraduate students. [Electronic Article]. *Journal of Education for Business, 86*(2), 64-70.

Cook, D. A., Levinson, A. J., Garside, S., Dupras, D. M., Erwin, P. J., & Montori, V. M. (2008). Internet-based learning in the health professions. [Electronic Article]. *JAMA: The Journal of the American Medical Association, 300*(10), 1181-1196.

Davies, R. S., Howell, S. L., & Petrie, J. A. (2010). A review of trends in distance education scholarship at research universities in North America, 1998-2007. [Electronic Article]. *International Review of Research in Open and Distance Learning, 11*(3), 42-56.

Dolinsky, B., & Kelley, J. M. (2010). For better or for worse: Using an objective program assessment measure to enhance an undergraduate psychology program. [Electronic Article]. *Teaching of Psychology, 37*(4), 252-256.

Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Englewood Cliffs, N.J.: Prentice Hall.

Educational Testing Service. (2010). Reliability coefficients and standard errors of measurement. Retrieved April 13, 2011, from http://www.ets.org/Media/Tests/MFT/pdf/mft_reliability_sem.pdf

Educational Testing Service. (2011a). 2011 major field test comparative guide. Retrieved March 11, 2011 from http://www.ets.org/s/mft/pdf/2011/business4gmf.pdf

Educational Testing Service. (2011b). About the major field test. Retrieved April 13, 2011, from http://www.ets.org/mft/about

Eduventures Learning Collaborative for Higher Education. (2006). *Expanding demand for online higher education, part 1: Assessing consumer views & experiences*. Boston, MA: Eduventures, LLC.

Engeström, Y. (1987). Learning by expanding: An activity-theoretical approach to developmental research. Available from http://lchc.ucsd.edu/MCA/Paper/Engestrom/expanding/toc.htm

Erwin, T. D. (2005). Standardized testing and accountability: Finding the way and the will. In J. C. Burke (Ed.), *Achieving accountability in higher education*. San Francisco: Jossey-Bass.

Evans, N., & Henrichsen, L. (2008). Long-term strategic incrementalism: An approach and a model for bringing about change in higher education. [Electronic Article]. *Innovative Higher Education, 33*(2), 111-124.

Ewell, P. T. (2007). Accreditation in the hot seat [from the states column]. *Assessment Update, 19*, 11-13.

Finkelstein, M. J. (2000). *Dollars, distance, and online education : The new economics of college teaching and learning*. Phoenix, Ariz.: Oryx Press.

Firestone, W. A., Schorr, R. Y., & Monfils, L. F. (2004). *The ambiguity of teaching to the test: Standards, assessment, and educational reform*. Mahwah, N.J.:Lawrence Erlbaum.

Frazier, T. W., & Edmonds, C. L. (2002). Curriculum predictors of performance on the major field test in psychology ii. [Electronic Article]. *Journal of Instructional Psychology, 29*(1), 29-32.

Frey, M. C., & Detterman, D. K. (2004). Scholastic assessment or g? The relationship between the scholastic assessment test and general cognitive ability. [Electronic Article]. *Psychological Science, 15*(6), 373-378.

Gebara, T. T. (2010). Comparing a blended learning environment to a distance learning environment for teaching a learning and motivation strategies course. Available from http://rave.ohiolink.edu/etdc/view?acc_num=osu1274276353

Geiser, S., & Studley, R. (2002). Uc and the sat: Predictive validity and differential impact of the SAT I and SAT II at the University of California. [Electronic Article]. *Educational Assessment, 8*(1), 1-26.

Gobert, J. D., & Koedinger, K. R. (2011). *Using model-tracing to conduct performance assessment of students' inquiry skills within a microworld*. Evanston, I.L.: Society for Research on Educational Effectiveness.

Heberlein, T. A., & Baumgartner, R. (1978). Factors affecting response rates to mailed questionnaires: A quantitative analysis of the published literature. [Electronic Article]. *American Sociological Review, 43*(4), 447-462.

Hosch, B. J. (2010). *Time on test, student motivation, and performance on the collegiate learning assessment: Implications for institutional accountability*. Paper presented at the Assocation for Institutional Research Annual Forum. Retrieved March 16, 2011 from http://www.ccsu.edu/uploaded/departments/AdministrativeDepartments/Institutional_Research_and_Assessment/Research/20100601a.pdf

Hu, S. E. (2005). Beyond grade inflation: Grading problems in higher education. Ashe Higher Education Report, (Vol. 30, pp. 1-99): Jossey Bass.

Hutchins, E. (1995). *Cognition in the wild*. Cambridge, Mass.: MIT Press.

Iseli, M. R., Koenig, A. D., Lee, J. J., & Wainess, R. (2010). *Automatic assessment of complex task performance in games and simulations. Cresst report 775*: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Issroff, K., & Scanlon, E. (2002). Using technology in higher education: An activity theory perspective. [Electronic Article]. *Journal of Computer Assisted Learning, 18*(1), 77-83.

Jahng, N., Krug, D., & Zhang, Z. (2007). Student achievement in online education compared to face-to-face education. [Electronic Article]. *European Journal of Open, Distance and E-Learning* Retrieved March 15, 2010, from http://www.eurodl.org/materials/contrib/2007/Jahng_Krug_Zhang.htm

Jones, D. J., & Watson, B. C. (1990). "High risk" students and higher education: Future trends. Eric digest. Washington, DC: ASHE-ERIC Higher Education Reports, The George Washington University.

Karathanos, D. (1991). Outcome measures of collegiate business curricula: A comparison of two instruments. [Electronic Article]. *Journal of Education for Business, 67*(2), 100.

Kerres, M., & Witt, C. D. (2003). A didactical framework for the design of blended learning arrangements. [Electronic Article]. *Journal of Educational Media, 28*(2-3), 101-113.

King, G. (1986). How not to lie with staticistics: Avoiding common mistakes in quantitative political science. [Electronic Article]. *American Journal of Political Science, 30*, 666-687.

Klein, S., Benjamin, R., Shavelson, R., & Bolus, R. (2007). The collegiate learning assessment: Facts and fantasies. [Electronic Article]. *Evaluation Review, 31*(5), 415-439.

Koenig, R. J. (2009). Comparing effectiveness of undergraduate course delivery: A student perspective. [Electronic Article]. *Journal of College Teaching & Learning, 6*(2), 69-80.

Koschmann, T. D. (1996). *Cscl, theory and practice of an emerging paradigm*. Mahwah, N.J.: L. Erlbaum Associates.

Kozma, R. B. (1994). Will media influence learning? Reframing the debate. [Electronic Article]. *Educational Technology Research and Development, 42*(2), 7-19.

Kraiger, K., Ford, J. K., & Salas, E. (1993). Application of cognitive, skill-based, and affective theories of learning outcomes to new methods of training evaluation. [Electronic Article]. *Journal of Applied Psychology, 78*(2), 311-328.

Lawrence, K. E., Reed, K. L., & Locander, W. (2011). Experiencing and measuring the unteachable: Achieving AACSB learning assurance requirements in business ethics. [Electronic Article]. *Journal of Education for Business, 86*(2), 92-99.

Lee, V. S., Hyman, M. R., & Luginbuhl, G. (2007). The concept of readiness in the academic department: A case study of undergraduate education reform. [Electronic Article]. *Innovative Higher Education, 32*, 3-18.

Leont'ev, A. N. (1978). Activity, consciousness, and personality. Available from http://www.marxists.org/archive/leontev/works/1978/index.htm

Lou, Y., Bernard, R., & Abrami, P. (2006). Media and pedagogy in undergraduate distance education: A theory-based meta-analysis of empirical literature. [Electronic Article]. *Educational Technology Research and Development, 54*(2), 141-176.

Lowenthal, P., & Wilson, B. (2010). Labels do matter! A critique of aect's redefinition of the field. [Electronic Article]. *TechTrends, 54*(1), 38-46.

Machtmes, K., & Asher, J. W. (2000). A meta-analysis of the effectiveness of telecourses in distance education. [Electronic Article]. *American Journal of Distance Education, 14*(1), 27 - 46.

Mansfield, E. R., & Helms, B. P. (1982). Detecting multicollinearity. [Electronic Article]. *The American Statistician, 36*(3), 158-160.

Martell, K. (2007). Assessing student learning: Are business schools making the grade? [Electronic Article]. *Journal of Education for Business, 82*(4), 189-195.

Mason, P. M., Coleman, B. J., Steagall, J. W., Gallo, A. A., & Fabritius, M. M. (2010). The use of the ETS major field test for assurance of business content learning: Assurance of waste? [Electronic Article]. *Journal of Education for Business, 86*(2), 71-77.

McLuhan, M. (1964). *Understanding media : The extensions of man* (2d ed.). New York: New American Library.

Mirchandani, D., Lynch, R., & Hamilton, D. (2001). Using the ETS major field test in business: Implications for assessment. [Electronic Article]. *Journal of Education for Business, 77*(1), 51.

Moody, J. (2004). Distance education: Why are the attrition rates so high? *Quarterly Review of Distance Education, 5*(3), 205-210.

Moore, J. L., Dickson-Deane, C., & Galyen, K. (2011). E-learning, online learning, and distance learning environments: Are they the same? [Electronic Article]. *The Internet and Higher Education, 14*(2), 129-135.

Nagel, D. (2009). Most college students to take classes online by 2014. *Campus Technology*. Retrieved January 30, 2011 from http://campustechnology.com/articles/2009/10/28/most-college-students-to-take-classes-online-by-2014.aspx

National Center for Education Statistics. Nontraditional undergraduates / definitions and data. Retrieved January 6, 2012, from http://nces.ed.gov/pubs/web/97578e.asp

National center for education statistics college navigator. (2011). Retrieved January 26, 2011, from http://nces.ed.gov/collegenavigator/?q=peirce+college&s=all&id=214883

Nelson, B. C., Erlandson, B., & Denham, A. (2011). Global channels of evidence for learning and assessment in complex game environments. [Electronic Article]. *British Journal of Educational Technology, 42*(1), 88-100.

Newman, S. E. (1957). Student vs. Instructor design of study method. [Electronic article]. *Journal of Educational Psychology October, 48*(6), 328-333.

Oliver, M., & Trigwell, K. (2005). Can 'blended learning' be redeemed? [Electronic Article]. *E-Learning, 2*(1), 17-26.

Orlikowski, W. J. (1992). The duality of technology: Rethinking the concept of technology in organizations. [Electronic Article]. *Organization Science, 3*(3), 398-427.

Priebe, L. C., Ross, T. L., & Low, K. W. (2008). Exploring the role of distance education in fostering equitable university access for first generation students: A phenomenological survey. [Electronic Article]. *International Review of Research in Open and Distance Learning, 9*(1), 1-12.

Rook, S. P., & Tanyel, F. I. (2009). Value-added assessment using the major field test in business. [Electronic Article]. *Academy of Educational Leadership Journal, 13*(3), 87-94.

Russel, D. (2002). Looking beyond the interface: Activity theory and distributed learning. In M. Lea & K. Nicoll (Eds.), *Distributed learning: Social and cultural approaches to practice* (pp. 64-82). London: Routlege/Falmer.

Russel, T. L. (2010). No significant differences frequently asked questions. Retrieved March 9, 2011, from http://www.nosignificantdifference.org/faq.asp

Russell, T. L. (1999). *The no significant difference phenomenon as reported in 355 research reports, summaries and papers*: North Carolina State University.

Saba, F. (1998). Is distance education comparable to "traditional" education? [Electronic Article]. *Distance Education Report, sample issue, 3*.

Salmon, G. (2004). *E-moderating: The key to online teaching and learning*. London:Routledge.

Schneider, B., Carnoy, M., Kilpatrick, J., Schmidt, W. H., & Shavelson, R. J. (2007). Estimating causal effects using experimental and observational designs. Retrieved April 2, 2011 from http://www.aera.net/publications/Default.aspx?menu_id=46&id=3360

Secretary's procedures and criteria for recognition of accrediting agencies, 34 C.F.R. § 602 (1987).

Seibold, K. (2007). *Employers' perceptions of online education.* Unpublished Ed.D. dissertation, Oklahoma State University, United States -- Oklahoma.

Shachar, M., & Neumann, Y. (2003). Differences between traditional and distance education academic performances: A meta-analytic approach. [Electronic

Article]. *The International Review of Research in Open and Distance Learning, 4*(2).

Shaftel, J., & Shaftel, T. L. (2007). Educational assessment and the AACSB. *Issues in Accounting Education, 22*(2), 215.

Shavelson, R. J. (2010). *Measuring college learning responsibly : Accountability in a new era*. Stanford, Calif.: Stanford University Press.

Simonson, M., Schlosser, C., & Hanson, D. (1999). Theory and distance education: A new discussion. [Electronic Article]. *American Journal of Distance Education, 13*(1), 60-75.

Sitzmann, T., Kraiger, K., Stewart, D., & Wisher, R. (2006). The comparative effectiveness of web-based and classroom instruction: A meta-analysis. [Electronic Article]. *Personnel Psychology, 59*(3), 623-664.

Smith, D. E., & Mitry, D. J. (2008). Investigation of higher education: The real costs and quality of online programs. [Electronic Article]. *Journal of Education for Business, 83*(3), 147(146).

Smith, P. L., & Dillon, C. L. (1999). Comparing distance learning and classroom learning: Conceptual considerations. [Electronic Article]. *American Journal of Distance Education, 13*(2), 6 - 23.

Spellings, M. (2006). *The spellings commission report : A test of leadership: Charting the future of u.S. Higher education*. Washington, D.C.: U.S. Department of Education.

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, Mass.: MIT Press.

Stoloff, M. L., & Feeney, K. J. (2002). The major field test as an assessment tool for an undergraduate psychology program. [Electronic Article]. *Teaching of Psychology, 29*(2), 92-98.

Suchman, L. A. (1987). *Plans and situated actions : The problem of human-machine communication*. Cambridge ; New York: Cambridge University Press.

Suskie, L. A. (2009). *Assessing student learning : A common sense guide* (2nd ed.). San Francisco: Jossey-Bass.

Suskie, L. A. (n.d.). The role of published tests and assessments in higher education. Retrieved April 13, 2011, from http://www.msche.org/publications/Published%20Instruments%20in%20Higher%20Education.doc

Szafran, R. F. (1996). The reliability and validity of the "major field test" in measuring the learning of sociology. [Electronic Article]. *Teaching Sociology, 24*(1), 92-96.

Tamim, R. M., Bernard, R. M., Borokhovski, E., Abrami, P. C., & Schmid, R. F. (2011). What forty years of research says about the impact of technology on learning: A second-order meta-analysis and validation study. [Electronic Version]. *Review of Educational Research, 81*(8), 1-25.

Taylor, J. C. (2007). 5th generation distance education: A sustainable approach to development. Retrieved April 25, 2011, from http://www.usq.edu.au/users/taylorj/publications_presentations/GDLN-APA%20Mk%20II.pps

Thompson, K. R. (2004). A conversation with Milton Blood: The new AACSB standards. [Electronic Article]. *Academy of Management Learning & Education, 3*(4), 429-439.

U.S. Department of Education Office of Planning Evaluation and Policy Development. (2010). *Evaluation of evidence-based practices in online learning: A meta-analysis and review of online learning studies*. Washington, D.C.: US Department of Education.

Ungerleider, C., & Burns, T. (2003). A systematic review of the effectiveness and efficiency of networked ICT in education. Retrieved March 15, 2011, from http://www.lnt.ca/technology/ict/SystematicReview.pdf

Vygotsky, L. S. (1978). *Mind in society : The development of higher psychological processes*. Cambridge: Harvard University Press.

Vygotsky, L. S., Van der Veer, R., & Valsiner, J. (1994). *The vygotsky reader*. Oxford, UK ; Cambridge, Mass.: Blackwell.

Wellman, B., Quan-Haase, A., Boase, J., & Chen, W. (2003). The social affordances of the internet for networked individualism. [Electronic Article]. *Journal of Computer-Mediated Communication, 8*(3).

Wenger, E. (1998). *Communities of practice : Learning, meaning, and identity*. Cambridge, U.K.: Cambridge University Press.

Williams, S. L. (2006). The effectiveness of distance education in allied health science programs: A meta-analysis of outcomes. [Electronic Article]. *American Journal of Distance Education, 20*(3), 127 - 141.

Winner, L. (1986). Do artifacts have politics? In L. Winner (Ed.), *The whale and the reactor: A search for limits in an age of high technology* (pp. 19-39). Chicago: University of Chicago Press.

Xu, D., & Jaggars, S. S. (2011a). The effectiveness of distance education across virginia's community colleges: Evidence from introductory college-level math and english courses. [Electronic Article]. *Educational Evaluation and Policy Analysis, 33*(3), 360-377.

Xu, D., & Jaggars, S. S. (2011b). *Online and hybrid course enrollment and performance in Washington State community and technical colleges. CCRC working paper no. 31.* New York, N.Y.: Community College Research Center.

Zwick, R., & Sklar, J. C. (2005). Predicting college grades and degree completion using high school grades and sat scores: The role of student ethnicity and first language. [Electronic Article]. *American Educational Research Journal, 42*(3), 439-464.

APPENDIX. UNIVARIATE CORRELATIONS

		MFT	Pct. OL	Transfer Credits	GE GPA	Major GPA	Age at test
MFT	Pearson Correlation	1	.281**	.205**	.394**	.419**	.059
	Sig. (2-tailed)		< .001	< .001	< .001	< .001	.090
	N	817	817	817	813	817	817
Pct. OL	Pearson Correlation	.281**	1	.308**	-.018	-.030	-.153**
	Sig. (2-tailed)	< .001		< .001	.605	.386	< .001
	N	817	817	817	813	817	817
Transfer Credits	Pearson Correlation	.205**	.308**	1	.015	.102**	-.033
	Sig. (2-tailed)	< .001	< .001		.678	.003	.346
	N	817	817	817	813	817	817
GE GPA	Pearson Correlation	.394**	-.018	.015	1	.844**	.178**
	Sig. (2-tailed)	< .001	.605	.678		< .001	< .001
	N	813	813	813	813	813	813
Major GPA	Pearson Correlation	.419**	-.030	.102**	.844**	1	.209**
	Sig. (2-tailed)	< .001	.386	.003	< .001		< .001
	N	817	817	817	813	817	817
Age at test	Pearson Correlation	.059	-.153**	-.033	.178**	.209**	1
	Sig. (2-tailed)	.090	< .001	.346	< .001	< .001	
	N	817	817	817	813	817	817
male	Pearson Correlation	.180**	-.010	.085*	.086*	.079*	-.056
	Sig. (2-tailed)	< .001	.765	.015	.015	.023	.110
	N	817	817	817	813	817	817
MAJ-ACC	Pearson Correlation	-.020	.007	.049	-.063	-.099**	-.068
	Sig. (2-tailed)	.567	.845	.163	.072	.005	.052
	N	817	817	817	813	817	817
MAJ-BSL	Pearson Correlation	.071*	.073*	-.050	.042	.030	.044
	Sig. (2-tailed)	.043	.036	.153	.235	.396	.213
	N	817	817	817	813	817	817
MAJ-HRM	Pearson Correlation	-.025	.094**	.009	.001	-.009	-.038

	Sig. (2-tailed)	.483	.007	.797	.976	.807	.275
	N	817	817	817	813	817	817
MAJ-MGT	Pearson Correlation	-.042	-.128**	-.062	.081*	.097**	.123**
	Sig. (2-tailed)	.234	< .001	.078	.022	.006	< .001
	N	817	817	817	813	817	817
MAJ-MKT	Pearson Correlation	-.024	.014	-.003	-.108**	-.064	-.175**
	Sig. (2-tailed)	.492	.685	.929	.002	.068	< .001
	N	817	817	817	813	817	817
MAJ-OTH	Pearson Correlation	.131**	.166**	.124**	.012	.009	.027
	Sig. (2-tailed)	< .001	< .001	< .001	.732	.788	.434
	N	817	817	817	813	817	817
African American	Pearson Correlation	-.282**	-.140**	-.044	-.315**	-.338**	.081*
	Sig. (2-tailed)	< .001	< .001	.206	< .001	< .001	.021
	N	817	817	817	813	817	817
Hispanic	Pearson Correlation	-.073*	-.058	-.080*	.004	-.020	-.012
	Sig. (2-tailed)	.036	.098	.022	.900	.575	.733
	N	817	817	817	813	817	817
Caucasian	Pearson Correlation	.298**	.140**	.037	.342**	.366**	-.010
	Sig. (2-tailed)	< .001	< .001	.289	< .001	< .001	.775
	N	817	817	817	813	817	817
Unknown	Pearson Correlation	.122**	.182**	.063	.010	.044	-.042
	Sig. (2-tailed)	< .001	< .001	.072	.770	.205	.230
	N	817	817	817	813	817	817
Other	Pearson Correlation	-.030	-.108**	.041	-.008	-.016	-.121**
	Sig. (2-tailed)	.389	.002	.247	.813	.643	.001
	N	817	817	817	813	817	817

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).