NOTE TO USERS

DISSERTATION

GENERAL EDUCATION GAINS FOR GRADUATES OF COLORADO'S LARGEST COMMUNITY COLLEGE

Submitted by

Laura M. Jensen

School of Education

In partial fulfillment of the requirements for the Degree of Doctor of Philosophy

Colorado State University

Fort Collins, Colorado

Spring 2005

COLORADO STATE UNIVERSITY

January 31, 2005

WE HEREBY RECOMMEND THAT THE DISSERTATION PREPARED UNDER OUR SUPERVISION BY LAURA M. JENSEN ENTITLED GENERAL EDUCATION GAINS FOR GRADUATES OF COLORADO'S LARGEST COMMUNITY COLLEGE BE ACCEPTED AS FULFILLING IN PART THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY.

Committee on Graduate Work

Gregory P. Smith

Randall Swaim

Leonard Albright

Jerry A. Magnan

Advisor

Department Head/Director

ABSTRACT OF DISSERTATION

GENERAL EDUCATION GAINS FOR GRADUATES

OF COLORADO'S LARGEST COMMUNITY COLLEGE

The purpose of the study was to present the general education relative gains at Front Range Community College (FRCC), utilizing the link between the American College Test Assessment (ACT) and the Collegiate Assessment of Academic Proficiency (CAAP). The specific research questions and corresponding results were as follows:

1. Do FRCC graduates' ACT scores differ from the reference group? Graduates' ACT scores on each domain were significantly higher than the reference group but the practical significance of the findings was marginal, as exhibited by small effect sizes.
2. Do FRCC graduates' CAAP scores differ from the reference group? Graduates' CAAP scores on each domain were significantly higher than the reference group and displayed moderate effect sizes.
3. Do FRCC graduates' gains differ from the reference group? Graduates' gains differed significantly from the reference group on every domain except writing. A lower proportion of graduates demonstrated lower than expected gains while a higher proportion of graduates demonstrated higher than expected gains.
4. Are FRCC graduates' gains equal across general education domain? Graduates' gains differed significantly by domain. Results indicated reading gains were significantly higher than gains on other domains.

5. Are FRCC graduates' gains equal across campuses? Graduates' gains did not differ significantly by campus with one exception; males at the Larimer campus had higher gains than females at the Westminster campus.
6. Are FRCC graduates' gains equal across degree granted? Graduates' gains differed significantly by degree on the math domain only. A.S. graduates had significantly higher math gains than all other degrees, and A.A. graduates had significantly higher gains than A.A.S. graduates. There were also significant differences on the math domain by gender and minority status.

This study incorporated a nationally-standardized methodology into an existing assessment framework to illustrate graduates' general education relative gains and yields, institutionally specific results and implications as well as more generalizable implications for possible incorporation into state-wide accountability initiatives.

Laura M. Jensen
School of Education
Colorado State University
Fort Collins, Colorado 80523
Spring 2005

ACKNOWLEDGEMENTS

The researcher would like to acknowledge with great appreciation those individuals who have contributed greatly to the completion of this research.

Dr. George Morgan, my committee chairman and advisor, who guided me through the academic process and encouraged me to explore my passion for research. Thank you.

My committee members, Dr. Lynn Albright, Dr. Greg Smith, and Dr. Randy Swaim who gave of their time and expertise to ensure my best effort. Thank you.

The Director of Institutional Research at Front Range Community College, Patricia Meade M.A., whose vision and leadership inspire me. I am grateful to you for your support and encouragement.

ACT Research Associate, Matt Harmston, for his assistance and support in the research design and data compilation. Many thanks.

Thank you to my family who has served as a never-ending source of motivation, and who has always encouraged me to believe in myself and dare to dream.

My deepest appreciation goes to my husband, Greg Jensen, whose support, encouragement, and love have carried and strengthened me. You are a blessing in my life that I am grateful for every day. I love you.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF FIGURES

Chapter I

INTRODUCTION

The discussion of assessment in higher education dates back to the early 1900s. As early as 1924, there were complaints about the need for assessment measures that could "register the difference between the ideals of college education and the actual gains made by students" (Sims, 1992, p. 24). The enactment of the Servicemen's Readjustment Act of 1944, better known as the GI Bill, and the higher education enrollment boom of the 1950s and 60s combined to further push the assessment agenda. While the agenda began to build support in the 1970s and gained momentum in the 1980s, nationally it was "the 1990s [that] clearly [witnessed] a change in policy makers' attitudes toward higher education. The general tone [was] one of less voluntary institutional improvement and more mandated public accountability" (Gaither, Nedwek, & Neal, 1994, p.38). In large part this change occurred because of an economic recession and claims of widespread inefficiency and ineffectiveness in higher education across the nation.

Colorado was no different. Beginning in 1985, with the passage of House Bill 1187, mandated accountability arrived. This legislation required institutions to submit assessment and accountability plans to the Colorado Commission on Higher Education (CCHE) but provided institutions autonomy with respect to specific methodologies. As Ewell (2002) noted, "despite HB1187's directive language and strong penalties, the essence of Colorado's assessment approach [was] decentralized. No state-wide instruments [were] mandated and institutions [were] given considerable discretion to define their own assessment approaches"

(p.165). The result was a tremendous amount of paperwork and stacks of institutional reports that may have gone largely unread.

Around this same time, the Colorado Community College and Occupational Education System (CCCOES) followed the lead of other governing boards and suggested the use of the Core Indicators for Effectiveness for Community Colleges published by the American Association of Community Colleges. Extending beyond the federal requirements of the Carl Perkins Applied Technology and Vocational Act, the indicators were intended to represent the breadth of the community college mission and incorporate indicators of student goal attainment, fall-to-fall retention, completion rates, employment rates, employer survey results, transfer rates, performance after transfer, college-level success for basic skills students, participation rates, demonstration of literacy skills, demonstration of citizenship skills, assessment of programs and services, and responsiveness to the community. While the CCCOES' intent was to present comparable data that were relative to the community college mission, stakeholder groups were seemingly not satisfied with the resulting information.

This dissatisfaction was the impetus for Colorado's first prescribed performance indicator system, which was created in 1996 with the repeal of HB 1187 and the passage of HB 1219 which directed the CCHE to develop a set of quality indicators. The resulting Quality Indicator System (QIS) was an attempt to mandate and prescribe exactly what would be measured so that institutions would submit comparable data that could be understood easily. It was modeled after many other performance indicator systems across the nation and was not specific to the community college mission. The QIS was typical in that it incorporated input and output information such as funding per full-time equivalent (FTE),

class size, graduation rates, employment rates and so forth, but failed to include learning outcomes for general education.

Currently, there is a shift taking place nationally that will extend beyond outputs and bring learning outcomes into the forefront of accountability and performance reporting. This evolution will incorporate, to a much larger degree than in the past, the measurement of demonstrable gains. Colorado is part of this shift. With the passage in May 2004 of Senate Bill 189, the College Opportunity Fund Act, the landscape of accountability in Colorado is evolving once again. "[We are entering] a new era for measuring the performance of higher education" (CCHE, 2004). Starting in July 2005, public postsecondary institutions in Colorado will begin operating under performance contracts with the CCHE. These performance contracts are an attempt to address some of the shortcomings of the QIS including the lack of general education learning outcomes as well as ease the tensions between centralization and institutional discretion by incorporating some degree of mission specificity. The intent is to create performance indicators that measure attainment of state-wide goals while also incorporating indicators specific to the role and mission of individual institutions. State-wide conversations among legislators, civic leaders, higher education administrations (including institutional researchers) and accountability experts were kicked off at the Colorado Higher Education Summit on Accountability and Performance in June 2004. Participants discussed what has been or can be learned from previous indicator systems and explored best practices for indicator systems including appropriate standards and methodologies (CCHE, 2004).

In his opening remarks at the Summit, the CCHE Executive Director, Rick O'Donnell, stressed three ideals for the new performance indicator contracts. First, the

indicators should focus on the knowledge and research outcomes an institution generates. In order to accomplish this, institutions should be able to measure gains/value-added. Second, new indicators should be "transparent" (CCHE, 2004). They should be able to make a point to a diverse group of stakeholders including but not limited to taxpayers, students, parents, and businesses, as well as communities the institutions serve. Third, indicators should portray high expectations. The idea of the performance contracts is to help raise the level of higher education quality in Colorado to a point where the institutions are national standards of excellence.

The current study addresses these ideals for performance contracts as they relate to Colorado community colleges. A distinction for community colleges is important because "[their] missions are typically much broader [than four-year institutions], including career and occupational programs, remedial and developmental course work, and various other educational offerings in addition to traditional liberal arts and sciences transfer programs" (Seybert, 2002). While a comprehensive performance contract for the Colorado Community College System (CCCS) should address each component of this mission, the current study is designed to focus only on the issue of graduates' general education gains. It utilizes general education assessment scores at admission, exit scores at graduation, and calculated gain levels to quantify general education program efficacy.

The study was completed at Front Range Community College (FRCC) for three reasons. First, FRCC is a multi-campus, two-year, public, postsecondary institution serving both rural and suburban populations. It is, therefore, representative of the community college system. Second, with an unduplicated student headcount of 23,926 and 9,874.38 student FTE generated during the 2004 academic year, it is the largest community college in

Colorado. Third, it is the only institution in the Colorado Community College System (CCCS) that utilizes any kind of standardized, objective, direct measure of general education when a student completes a degree. The combination of the multiple locations, institutional size, and unique assessment policy make it an ideal setting in which to pilot research to directly measure student learning in general education.

Rationale for the Study

The creation of performance contracts presents an opportunity for governing boards and institutions in Colorado to negotiate with the state performance measures that reflect their role and mission. The study originated from the need for accountability and rests on the triangular relationship among the issues of feedback, measurement, and policy appropriate methodology as they relate to the efficacy of general education in Colorado community colleges.

Feedback: The feedback loop of assessment is important to institutions for a variety of reasons including improvement, planning, accountability, public perception, political lobbying, and accreditation. As the Higher Learning Commission (the accrediting body for Colorado community colleges) states, institutional self-assessment should provide for institutional evaluation, assure accountability in the use of federal funds allocated to the institutions, and promote, strengthen, and assure the operation of quality educational programs (2000). The feedback process, therefore, requires community colleges to assess not only their vocational programs (as federally mandated by the Carl Perkins Applied Technology and Vocational Act and the Workforce Investment Act) but also their general education program across the curriculum. If the efficacy of a general education program is

defined as successful student learning, then an institution's feedback loop should include a direct measure of graduates' general education gains.

Measurement: There are a variety of approaches to assessing student learning in general education programs including, but not limited to, standardized objective tests, locally designed tests, surveys, capstone courses, portfolios and focus groups. The CCCS utilizes nationally standardized objective tests to assess the basic skills and general education of incoming students. This type of assessment is used for numerous reasons. For instance, issues of reliability and validity have already been addressed on nationally-normed instruments, and administration is relatively quick and fiscally efficient when compared to other approaches. Additionally, the widespread use of nationally-developed instruments allows for comparisons to similar systems, institutions, and students. This provides a basis of interpretation by setting a benchmark against which performance can be measured. For these same reasons, a state-wide performance indicator of general education gains may need to employ standardized assessment measures.

Policy appropriate methodology: There are three pieces of legislation that could work in conjunction toward the assessment of general education outcomes in the CCCS. Each piece links together to pave the way for the efficient and effective measurement of general education gains.

1. Colorado Regulatory Statute 22-7-409: "Beginning in the spring semester 2001, and each spring semester thereafter, all students enrolled in the eleventh grade in public schools throughout the state shall be required to take a standardized, curriculum-based, achievement, college entrance examination selected by the department, administered throughout the United States, and relied upon by institutions of higher education that, at a

minimum, tests in the areas of reading, writing, mathematics, and science " (http://198.187.128.12/colorado/lpext.dll?f=templates&fn=fs-main.htm&2.0). The state chose to utilize the American College Test (ACT) assessment to fulfill the requirements of this statute. In 2003, the test was administered at the expense of the state to 44,485 eleventh grade students (http://www.cde.state.co.us/cdeassess/results/2003/ACT_StSum.xls). Using these same scores as the starting point for assessing postsecondary general education gains is fiscally efficient.

2. Colorado Regulatory Statute 23-1-113: This statute gives a directive to the CCHE regarding the assessment and tracking of basic skills. To comply with the statute, CCCS board policy 9-41 provides for the basic skills assessment of first-time, degree- seeking freshman in reading, writing, and mathematics and attempts to create comparability across institutions by requiring the utilization of specific assessment instruments. The policy states that incoming students should be asked to present their American College Test (ACT) assessment scores when applying for admission to a state-supported community college. Those students whose scores do not meet state-wide minimum requirements for college-level placement, are more than two years old, or are unavailable are then required to take the Accuplacer (a standardized objective basic skills and general education assessment distributed by the College Board organization). This represents a policy shift at the system level from institutionally-chosen placement assessments to common methodology and instrumentation across institutions. Any general education gains analysis should attempt to work within the framework of this existing policy.

3. Senate Bill 04-189: In May 2004, Colorado became the first state in the nation to move to stipend funding for higher education through the enactment of Senate Bill 04-189, the College Opportunity Fund Act. As previously stated, the passage of this bill not only changes the funding process for higher education, it also replaces the current indicator system with performance contracts. Exactly what this means is currently being negotiated; however, from discussions at the Colorado Commission on Higher Education Accountability and Performance Summit (June 2004), it seems likely that the CCCS will negotiate one performance contract for all community colleges in the system. Any general education gains analysis methodology should, therefore, be generalizable across all 13 colleges in the system.

Statement of the Problems

Just as the rationale for the proposed research originates from the need for accountability and rests on the triangular relationship among the issues of feedback, measurement, and policy appropriate methodology, so too do the specific problems that the study is designed to address.

Feedback: Currently, most community colleges in Colorado do not directly evaluate the general education gains made by graduates. This leaves a significant gap in the feedback loop for developing policies and practices relating to a variety of issues such as remedial education, general education curricula, transferability and so forth. The current study employs a method for closing this gap in the feedback loop for Front Range Community College by describing graduates' general education gains relative to their peers and conducting further analyses to identify gain patterns across educational domains, campuses, and degrees granted.

Measurement: While community colleges are mandated to respond to federal accountability requirements for vocational education through the Carl Perkins Applied Technology and Vocational Act and the Workforce Investment Act, accountability requirements for the rest of the curriculum are left for states or institutions to define and implement. Most preliminary attempts utilized the 14 indicators suggested by the American Association of Community Colleges (AACC) as a foundation for creating performance reports. This included input and output information but did not directly measure student learning outcomes in general education. The current methodology provides a direct measure of relative learning outcomes by extending the current general education assessment at FRCC with a nationally-normed gains analysis.

Policy appropriate methodology: There is currently no methodology within the CCCS that measures general education gains of graduates, let alone one that incorporates existing policies in order to be as efficient as possible with dwindling resources. While the proposed study is specific to FRCC, the largest college in the system, the methodology incorporates state-wide assessment policy and instrumentation, thereby allowing for possible future generalizability to each of the other colleges within the system.

Research Objectives

The current research objectives address the previously stated problems as follows:

1. Close the feedback loop at Front Range Community College (FRCC) with a direct measure of general education relative gains in writing, reading, mathematics, and science skills.

2. Employ a nationally validated methodology that can incorporate CCCS assessment policy to directly measure general education relative gains in writing, reading, mathematics, and science skills.

Research Questions

There are six research questions addressed by the current research.

1. Do FRCC graduates' ACT scores differ from the reference group?
2. Do FRCC graduates' CAAP scores differ from the reference group?
3. Do FRCC graduates' gains differ from the reference group?
4. Are FRCC graduates' gains equal across general education domain?
5. Are FRCC graduates' gains equal across campuses?
6. Are FRCC graduates' gains equal across degree granted?

Delimitations

The study is limited to an institutional-level analysis of four general education domains. It is not intended to be a comprehensive analysis of all general education domains. Additionally, data are from FRCC graduates with no other community colleges represented. Furthermore, the data are limited to four academic years (2001, 2002, 2003, and 2004).

Assumptions

It is assumed that the data in the Student Information System (SIS), maintained by the CCCS, are accurate and complete and that students gave their best effort on all assessments. Furthermore, it is assumed that the data provided by ACT are accurate and complete.

Operational Definitions

Academic Year: three consecutive semesters beginning with summer and crossing over two calendar years (ex: summer 2003, fall 2003, and spring 2004 constitute the 2004 academic year)

American College Testing (ACT): a standardized, objective instrument designed to assess high school students' general education development and their ability to complete college-level work. The test covers four skill areas: English, mathematics, reading, and science (http://www.act.org/aap)

Associate Degree: degree granted for the successful completion of a sub-baccalaureate program of study, usually requiring at least 2 years (or equivalent) of full-time college-level study (National Center for Education Statistics (NCES), Digest of Education Statistics, 2002)

Associate of Applied Science Degree (A.A.S.): two-year vocational degree to prepare students for employment

Associate of Arts Degree (A.A.): two-year (or equivalent) transfer degree for students who plan to major in a liberal arts program at the baccalaureate level

Associate of General Studies Degree (A.G.S.): two-year (or equivalent) degree for students who plan to major in a broad program of study including both vocational and academic courses

Associate of Science Degree (A.S.): two-year (or equivalent) transfer degree for students who plan to major in a science program at the baccalaureate level

Basic Skills Assessment Policy: Colorado Community College System (CCCS) board policy 9-41 providing for the basic skills assessment of first-time, degree-seeking freshmen

in reading, writing, and mathematics in accordance with the Colorado Commission on Higher Education (CCHE) state-wide remedial education policy (C.R.S. 23-1-113.3)

Collegiate Assessment of Academic Proficiency (CAAP): norm-referenced postsecondary examination of writing, mathematics, reading, science reasoning, and critical thinking skills

Common Data Set (CDS): standardized data elements and operational definitions utilized by the United States Department of Education to help ensure comparable data across higher education institutions

Community College: two-year postsecondary institution offering transfer and vocational degrees (intended to be completed in four semesters of full-time attendance) as well as vocational certificates and personal interest courses

Degree: award conferred by a college, university, or other postsecondary education institution as official recognition for the successful completion of a program of study (Common Data Set, 2002)

Degree-Seeking Student: student enrolled in credit courses and recognized by the institution as seeking a degree or formal award in academic or vocational programs

Direct Measure of Learning: requires students to demonstrate their knowledge and skills

First-Time Student: incoming student who has no prior earned credits from any higher education institution since high school graduation or equivalent certification

Gain Level: level of gain between a student's ACT scores and CAAP scores. Gain levels are categorized as below expected, expected, or above expected based on the distribution of CAAP deciles conditional on ACT deciles.

General Education: encompasses proficiencies and breadth of knowledge across multiple domains: reading, writing, math, science, ethics, literacy, citizenship, and critical thinking

Graduate: student who is granted an associate degree

Indirect Measure of Learning: requires students to reflect on their learning rather than demonstrate it

Open Admission Policy: admission policy under which virtually all secondary school graduates or students with General Equivalency Diplomas are admitted without regard to academic record, test scores, or other qualifications (Common Data Set, 2003)

Public Institution: educational institution whose programs and activities are operated by publicly elected or appointed school officials and which is supported primarily by public funds (Common Data Set, 2003)

Reference Group: national sample of public two-year postsecondary students (n= 23,999) who have taken both the ACT assessment and the CAAP test

Significance of the Research

Results of the current research would offer a direct measure of general education gains and gain patterns for Front Range Community College relative to graduates of other two-year public colleges who took the ACT and CAAP. This provides institutionally-specific significance. While the current institutional-level general education analysis at FRCC does answer the question, "How do FRCC graduates compare with graduates from similar institutions across the nation with regards to reading, writing, mathematics, and science skill levels," it does not address the fundamental question, "Relative to their peers, do

FRCC graduates demonstrate general education gains?" Nor does it address, "How do those gains compare to similar students, at admission, across the nation?"

Furthermore, the current research methodology has system-wide significance. All Colorado community colleges are under pressure to respond to increasing demands for fiscal accountability and institutional effectiveness with regards to general education. Although the upcoming performance contracts are yet to be defined, general education will be a topic of discussion. In testimony to the National Commission on Accountability in Higher Education, Nancy Shulock (2004) emphasized that the future of accountability should include "valid, aggregate measures of college-level learning such as the Collegiate Assessment of Academic Proficiency " (p.3). The current study utilizes this nationally-validated instrument and the corresponding gain levels compared to ACT scores to respond to these demands while incorporating already existing general education and basic skills assessment policies of Colorado K-12 and postsecondary education systems.

Chapter II

LITERATURE REVIEW

Introduction

Across the nation, higher education is facing ever-increasing demands for accountability. While there is debate about whether community colleges or their university counterparts have more difficulty meeting these demands (Banta, 1999), there is no doubt that the demands are here to stay and that institutions, systems, and states would do themselves a favor to take a proactive role in designing accountability measures that tell an honest and complete picture for stakeholders. This literature review examines assessment, accountability, and performance indicator systems in the community college sector of higher education. It then narrows the discussion to the Colorado Community College System (CCCS) and, more specifically, to the system's largest institution, Front Range Community College (FRCC). Additionally, the current study points to where accountability may be headed in Colorado and suggests the incorporation of an analysis of general education relative gains for FRCC that may have generalizability to the entire CCCS.

The Evolution of Accountability

The evolution of accountability in the community college sector parallels that of higher education in general. The higher education accountability movement of the 1990s began with the assessment movement of the 1970s and 80s. Although not entirely accurate, the distinction was that assessment, historically, was usually internally motivated and that accountability was externally mandated. The next evolution was from accountability to

performance reporting. For many institutions, this shifted the focus from reporting inputs and processes to reporting outputs and outcomes such as institutional performance and the performance of students and graduates. The most recent evolution resulted from an even stronger external push that tied funding allocations to reported performance. A description of the development and implementation of each of these evolutionary trends at the community college level is subsequently provided.

Assessment

"The emphasis in the 1980s was on the quality of higher education, especially undergraduate education. The student outcomes assessment movement was the response to this emphasis" (Serban, 1998, p. 22). Assessment should be closely related to an institution's mission and will therefore be defined slightly different across campuses. However, Palomba and Banta (1999) generally defined assessment as "the systematic collection, review, and use of information about educational programs undertaken for the purpose of improving student learning and development" (p. 4). As this definition implies, effective assessment can, therefore, be described as a continuous feedback loop. For community colleges, the mission specificity of outcomes assessment requires defining measurement for a multitude of educational components. An understanding of this requires consideration of each component of the community college mission.

Initially, public two-year colleges were established to train workers for expanding industries. It is still a mainstay of these institutions to respond to the needs of business and industry by providing relevant, quality vocational education for the workforce. One example of this is the partnership that community colleges have made with workforce centers to

provide a variety of training and employment services to assist individuals in successfully obtaining and retaining employment.

Additionally, community colleges are a primary resource for students requiring remedial education. As reported by the National Center for Education Statistics (NCES) (2004), in fall 2000 "remediation was more likely to be offered by public 2-year colleges (98 percent) than all other institutional types" (nces.ed.gov/surveys/peqis/publications/2004010/2.asp). Remedial education provides access to postsecondary education for students who are not prepared academically for college-level course work. Across the nation, there are an alarming number of high school graduates who are under-prepared for college-level course work. For example, according to the same NCES study, only 24 percent of 12th-graders nationally had proficient writing skills. These students, as well as some adult learners returning to school after extended absence from the classroom, may require these remedial services. Remediation prepares these students to be successful in higher education and is a responsibility that is increasingly falling to community colleges.

Furthermore, as Cohen and Brawer (2003) described, initially "educators wanted the universities to abandon their freshman and sophomore classes and relegate the function of teaching adolescents to a new set of institutions... All insisted that the universities would not become true research and professional development centers until they relinquished their lower-division preparatory work" (p. 6). Offering quality lower-division courses at community colleges not only appeased the university academics but also benefited the student by addressing issues of class size, location, affordability, and inclusiveness, especially for populations traditionally underrepresented in higher education. These lower-

division transfer courses are still fundamental to the community college mission. As reported by the National Center for Education Statistics (2003), "one-half of undergraduates start at a public 2-year institution with the intention of obtaining a bachelor's degree and about one-fourth of those who start with an associate's degree goal transfer to a 4-year institution within 6 years" (http://nces.ed.gov//programs/coe/2003/section3/indicator19.asp).

Moreover, community colleges and their service areas forge partnerships for sharing resources and promoting economic and cultural development. Examples of shared resources are joint-use libraries, visual and performing arts facilities, and training or conference facilities. Not only can these facilities be shared by K-12 and higher education, community groups and businesses may also utilize them.

For these reasons, Seybert (2002) summarized that community college missions are typically broader than four-year institutions. "In addition to traditional freshman- and sophomore-level course work, community colleges provide career training, occupational retraining, remedial and developmental course work, community and education programs, contract training for business and industry courses for special populations and a variety of other educational offerings" (p. 250). There are those who believe that the breadth of this mission complicates the assessment process. While this may be true, there are additional considerations as to why assessment at community colleges is distinct from other sectors of higher education.

Another consideration for community college assessment is the diversity of students who attend two-year institutions. As Seybert (2002) again summarized, "they are more diverse in terms of age, background, employment status, preparation, and educational objective than their four-year college and university counterparts" (p. 250). The NCES

(2004) *Condition of Education* report provided data on a variety of demographics for undergraduate students (nces.ed.gov/programs/coe). These data indicated that community college students are, on average, older than students at four-year schools. This means that many students are not coming to a community college immediately after graduating from high school. These students have life circumstances different from those of traditional age students and probably have many responsibilities competing for their resources of time, energy, and money. They may work, may have children, or may be changing careers, and school may not be the only focus of their busy lives. Another equally important point of diversity mentioned in the same study is that ethnic minorities represent a larger percentage of the student population at community colleges than they do at traditional four-year institutions. Yet another point of diversity, and arguably the one that most impacts student outcomes assessment, are the reasons that students attend two-year institutions. A student may have the goal of taking only one or two courses, obtaining a vocational certificate or an associate degree, or transferring to a four-year institution (either before or after graduation). These diverse student goals make assessing outcomes difficult because a successful outcome will mean something different depending on what the student's goals were initially. Irrespective of these challenges, there are examples of quality assessment across community colleges and, as Banta (1999) suggested, in some areas community colleges may be setting the standard for assessment in higher education.

An examination of a national sample categorized outcome assessments into three basic models based on the intended primary use of the data (Halpern, 1987). As Halpern delineated, the primary intended use of data might be either 1) the ability to gauge academic

progress (at either the individual or programmatic level), 2) to ensure basic competencies, or 3) to satisfy accountability mandates.

"Assessments of student outcomes are frequently used to provide information to individual students so that they can gauge their academic progress. When aggregated, the same information can be used to make programs (for example: academic major, general education, and certificate programs) more effective in meeting their goals and the overall mission of the institution" (Halpern, p. 7). These data are of use to both the individual student as well as the institution as a whole.

Then again, outcomes assessment data might be used primarily "to ensure basic competencies in all graduates" (Halpern, p. 7). Rising junior examinations or assessment of general education for graduates are both examples of this approach, which Halpern labeled the gateway model. As she points out, this model attempts to ensure a basic level of competency by collecting direct measurement of students' skills and knowledge. As one example of this, in 1983, Florida mandated the use of the College Level Academic Skills Test (CLAST) for all undergraduates completing their sophomore year (Van de Water, 1994).

Alternatively, data may be primarily used in accountability and performance budgeting and/or funding decisions. This model gained popularity in the early 1990s from disappointment with assessment plans that sacrificed comparability for institutional autonomy. What resulted from the other models of outcomes assessment were reports that held significance at the institutional or programmatic level but lacked comparability across institutions, systems, or states. Assessment for the primary purpose of accountability attempts to mediate this lack of comparability by utilizing performance indicators, a

methodology borrowed from the K-12 reform of recent years. Accountability outcomes assessment is, therefore, intended to provide readily understandable information for a broad group of stakeholders including students, parents, taxpayers, legislators, governing boards, and community leaders (Serban, 1998).

From Assessment to Accountability

As previous described, the assessment movement began to build support in the 1970s and 1980s, however, it was "the 1990s [that] clearly [witnessed] a change in policy makers' attitudes toward higher education. The general tone [was] one of less voluntary institutional improvement and more mandated public accountability" (Gaither, Nedwek, & Neal, 1994, p. 38). Nettles et al. (1997) provided a review of the research that supports this statement. In 1987, the Education Commission of States (ECS) and the State Higher Education Executive Officers (SHEEO) jointly published findings from a survey of all 50 states that indicated two-thirds of states had formal assessment policies but the general trend was toward institutional autonomy in design and implementation. Two years later, in 1989, these two groups teamed up with the American Association for Higher Education (AAHE) for a second survey with results that indicated assessment was more clearly defined at the state level and that states were taking a more active role in policy development. At that time, only eight states reported having no assessment policy. Six years later, in 1995, the National Center for Education Statistics (NCES) asked workshop participants to complete a pre-workshop inventory describing assessment in their state. Findings indicated a distinct shift from assessment for improvement to public accountability.

The underlying purpose of accountability systems is to improve higher education. However, "on the surface the purpose may appear to be narrowly instrumental: to identify and punish low performance, or to provide incentives or rewards for higher performance" (Lingenfelter, 2001, p.1). To avoid skimming the surface only, the design of accountability systems should be a collaborative effort among stakeholder groups (legislators, administrators, students, parents, teachers, business leaders, etc.). Resulting systems should primarily reflect what stakeholder groups find valuable about education and balance issues regarding the societal and economic impact of education, financial efficiency and increased effectiveness. Stakeholder discussions should operationalize accountability measures by specifying what will be measured, how it will be measured, and what data will need to be compiled. Issues of responsibility and obligation must also be addressed. The full burden of successful outcomes cannot fall on any one stakeholder. For example, the student cannot be held solely responsible for graduating within a certain number of credit hours because some of the responsibility for this must lie with the institution to offer the necessary courses. The responsibility is shared, as is the obligation of each stakeholder group to the others. The more clearly initial discussions can answer these questions, the more effective the resulting accountability system will be.

There are three traditional approaches to accountability systems, as identified by Lingenfelter (2001): compliance auditing, accreditation, and program review. "Compliance auditing is the precursor to other approaches of accountability and still plays an important role" (p. 5). Compliance auditing is a review process for institutions to ensure that federal and state guidelines are being followed.

Accreditation is another approach to accountability. There are seven regional accrediting organizations to which institutions may voluntarily belong: the Middle States Commission on Higher Education, the Commission of Institutions of Higher Education of the New England Association of Schools and Colleges, the Higher Learning Commission of the North Central Association of Colleges and Schools, the Northwest Commission on Colleges and Universities, the Commission on Colleges of the Southern Association of Colleges and Schools, the Western Association of Schools and Colleges: Accrediting Commission for Community and Junior Colleges, and the Western Association of Schools and Colleges: Senior Commission. "Accreditation is the recognition that an organization meets clearly defined criteria and that there are reasonable grounds for believing that it will continue to meet them. [It] provides both public certification of acceptable institutional quality and an opportunity and incentive for self-improvement" (Abt, 2004).

Program and policy reviews are another traditional approach to accountability. "In the last quarter of the 20th century, as the rapid rate of enrollment growth subsided in most states, state-wide agencies focused more attention on program quality and various public priorities" (Lingenfelter, 2001, p. 5). Agencies identified public priorities, such as increasing nursing graduates, and encouraged the reallocation of funds to support them. On the reverse side of this prioritization, programs with low enrollments and low demand in the workforce for graduates were encouraged to shut down. As part of this process, vocational programs must provide outcomes data about their students and graduates.

Lingenfelter (2001) also identified two emerging approaches to accountability: performance reporting and performance budgeting and funding. These approaches have become the next evolution of accountability across the nation and are the result of

increasingly tight budgets and frustrations with previous accountability systems (Banta & Borden, 1994, Burke, 1998, Ewell & Jones, 1994).

From Accountability to Performance Reporting

"The apparent inability of assessment to respond to the accountability concerns of state governments seems to have been one of the major reasons for the move to performance reporting in the late 1980s and early 1990s" (Serban, 1998, p. 23). The primary frustration with previous accountability attempts was the lack of comparability across institutions, systems, and states. The other major factors in the evolution of performance reporting were the increased complexity and size of higher education, rising costs in the face of dwindling financial resources, and concerns about improving the linkage between higher education and the rest of society (Ewell & Jones, 1994). By 2000, thirty states had implemented some sort of performance reporting initiative (Burke & Minassians, 2002). Initially, performance reporting was modeled after the performance report cards that resulted from the K-12 reform mandated by the No Child Left Behind Act. In K-12 education, this reform has increased accountability for results by mandating annual performance reporting at the school, district, and state level that provides information about outcomes and progress on a set of common indicators. This model also requires schools to complete the feedback loop of assessment by incorporating outcome data in decision making. Institutions not making adequate yearly progress are required to implement dramatic changes to the way the school is run. Postsecondary performance reporting followed this lead with only moderate success because it did not require common indicators or the incorporation of results in institutional decision

making. It was this frustration with the lack of internalization and commonality that gave way to the next evolution of accountability.

From Performance Reporting to Performance Budgeting/Funding

Understanding that money can usually influence behavior, states began to bring dollars into the performance reporting equation by creating performance budgeting and/or funding processes. As Burke (2002) reported, "linking budgeting and funding to performance emerged as a popular phenomenon in the middle '90 and by 2000, the practice had spread to three quarters of the states" (p. 19). While most states use the terms performance budgeting and performance funding synonymously, there are subtle differences. Performance funding directly ties resources to specific achievement of performance indicators. Performance budgeting allows for consideration of performance on indicators, but the funding link is indirect and at the discretion of the funding agency. The advantages and disadvantages are many. While performance funding is more apt to influence institutional change by rewarding success, it is also more apt to punish institutions with substandard performance due to circumstances out of their control. For instance, a rural community college may have substandard minority enrollment because of the lack of ethnic diversity in the service area population. This may not be something over which the institution has control. Additionally, some indicator systems have benchmarks that change every year in an attempt to encourage continuous improvement. Benchmarks that are moving targets are also out of institutions' control. Performance budgeting allows for more consideration of these types of extraneous variables. Furthermore, while the indicator commonality of performance funding allows for comparability across institutions, systems,

and states, it generally provides a very limited picture of institutional success by allowing only a handful of indicators. Performance budgeting allows for an increased number of indicators to be considered but may sacrifice some indicator commonality to allow institutions to develop a more complete picture of their performance.

The adoption of performance indicators has been met with some degree of resistance, regardless of whether the indicators are prescribed or not, or whether they are intended for reporting, budgeting, or funding. Of course, the most intense resistance has been to performance funding, especially when educational researchers have been excluded from discussions about the design of indicator systems. Burke (1998) delineated and refuted the major arguments for resisting performance based funding, many of which apply also to performance reporting and budgeting.

Argument 1 (Complexity): Central to this argument is that the goals of higher education are too numerous and complex for inclusion in a reasonable number of indicators. It is true that the mission of higher education is complex and reflects a variety of societal priorities. This is perhaps most true at the community college level. While this does make it difficult to paint a complete portrait with a handful of performance indicators, it does not preclude assessment of students, graduates, and institutions. Prescribed indicators reflect the societal values with the highest priority and are still of interest for the part of the picture they can represent.

Argument 2 (Diversity): The point of this argument is that the differences between community colleges, comprehensive universities, and research institutions cannot be accounted for in an indicator system. This argument expects rigidity in indicators that is not necessary. Indicators can be somewhat adjusted for each institutional type. For example,

graduation rates are typically measured at the six-year mark for four-year institutions and at the three-year mark for community colleges.

Argument 3 (Quality): The argument that the quality of higher education programs and services are too elusive for measurement is weak. There must be empirical evidence, not just anecdotes, to support claims of quality programming. Performance indicators are not an attempt to measure every aspect of quality but, instead, to measure those deemed most important by stakeholder groups.

Argument 4 (Funding): The main point to this argument is that "[performance funding] programs provide too little money to produce campus changes or allocate so much that they produce budget instability" (Burke, 1998, p. 86). It is true that the dollars tied to performance indicators must be an appropriate amount so as to encourage change without punishing too severely campuses with substandard performance. Most states have addressed this by linking performance indicators with additional funding, not cuts in general fund appropriations. As Burke clarified, "performance funding constitutes an added component and not a total substitute for budgeting based on workload driven by student enrollment, institutional missions, and program costs" (p. 86).

Argument 5 (Politics): This argument states that political landscapes change too often to support stable reporting systems. This same argument could be made of any political process and, yet, it does not bring government to a stand still. The stability of reporting systems comes not from political affiliation but from careful deliberations in design and implementation. There are many systems that have stood the test of time, such as those in Tennessee that serve as models for withstanding changes in political leadership.

Argument 6 (Cost): Another argument is the cost of collecting the data for reporting is a burden to institutions. While it is true that collection processes can be cumbersome initially, performance indicators measure those educational outcomes most valued by stakeholder groups, and these stakeholder groups would be surprised if institutions were not, in some way, already collecting these data. Additionally, indicator systems usually allow for a phase-in period during which data collection processes can be stabilized.

Argument 7 (Incompatibility): This argument postulates that externally motivated assessments, such as performance reporting, directly compete with internally-motivated assessment for the purpose of quality improvement. This rationale assumes that data cannot be used for the dual purpose of improvement and accountability. This assumption is a fallacy based largely in fear of retribution for data that indicate a need for improvement.

Argument 8 (Punishing the poor): This argument centers on the idea performance funding punishes institutions with already limited resources by taking additional funds. This can, again, be partially resolved with the use of discretionary funds as rewards for institutions exceeding standards instead of withholding funds for poorly-performing institutions. Additionally, systems can provide for institutions showing consistent improvement to the point that minimum standards are met.

A better understanding of performance reporting and performance budgeting and funding would go a long way toward easing concerns and countering these arguments.

Typology of Performance Indicators

Most broadly defined, performance indicators identify and articulate educational priorities for which achievement is measured and reported. They are an attempt to quantify,

with empirical, policy-relevant data, the functioning of an institution or system of higher education. As Banta and Borden (1994) described, contrasted with descriptive statistics and management information, performance indicators provide context in comparing and improving institutions because they are rooted in a goal-driven process.

Types of performance indicators include inputs, processes, outputs, and outcomes (Burke, 1998, Ewell & Jones, 1994, Richardson, 1994). Input indicators refer to resources received (human, financial, and physical) to support programs and activities. Common input indicators include tuition and fee dollars, general fund dollars, entering number of students, faculty credentials, ACT or SAT entrance scores, and student-to-faculty ratios. Inputs are usually representative of the most readily available data for institutions but fail to explain what the institution or system is doing and how well it is doing it. Richardson (1994) offers a rationale for why these indicators are often utilized. He stated, "the number of input measures included among state indicators would probably be even fewer if it were not for institutional pressures to include measures that demonstrate under-funding in comparison with peers in other states" (p. 135).

Process indicators describe the methods used to deliver services. Common process indicators include faculty workload, workforce training relationships, K-12 cooperative efforts, non-instructional and instructional expenditures, technology in the classroom, amount and type of distance education, assessment plans, course availability, program review, class size, continuing education, developmental services, economic development partnerships, faculty reviews, internships and cooperative education sites, courses at nontraditional times, and space utilization, among others. The data for these indicators are not usually as readily

available as those of input indicators and may also be derived through a variety of methods, making them less uniform across institutions than input indicators.

Output indicators involve the quantity of what is produced, while outcomes indicators refer to the quality of programs, activities, and services for students and society as a whole. Common output indicators include retention, graduation, and transfer rates (overall and minority), core course completion rates, and FTE (credit and non-credit) generated. Common outcomes indicators include job placement rates, graduates' salaries, licensure examination scores, satisfaction surveys (student, alumni, graduate, and employer), general education test scores, test scores for majors, remediation test scores, and program accreditations.

As Ewell and Jones (1994) stated, effective performance indicators, irrespective of type (input, process, output, or outcome) "must be more than simply a readily calculable statistic" (p. 28). They must bridge the gap between data and policy relevant information. Stakeholders are subject to making erroneous decisions if the data do not translate into useful information that is linked to policy. Ewell and Jones continued to describe the criteria for judging indicator effectiveness as follows.

Policy leverage addresses the extent to which any proposed indicator can identify deficiencies and influence corrective actions. "Some indicators merely offer information about the present state or condition of higher education without yielding much information about how to improve things; others provide direct guidance about what might be changed" (Ewell & Jones, 1994, p. 14). For example, with the national nursing shortage of the past few years, an indicator of the number of students wait-listed for nursing programs would merely provide information. If the indicator also calculated the number of clinical site

partnerships and identified actions that it would take to increase these partnerships, it would have more policy leverage.

Vulnerability addresses the extent to which any proposed indicator is susceptible to manipulation without real changes having occurred. Indicators must not be overly susceptible to manipulation (intended or otherwise). For example, a faculty workload indicator intended to measure how many hours are actually spent teaching can be manipulated when the regular and overload contracts are generated. If an instructor teaches three classes and is also chair of two committees, his/her workload is greater than regular full-time. If the regular contract is generated to reflect the three classes and an overload contract is generated for the committee work, the instructor appears to spend all of his/her regular hours in the classroom. If the regular contract is generated to reflect one class and two committees and the overload contract reflects the other two classes, the statistic changes entirely even when the work has not changed.

Balance of perspective refers to the degree to which indicators accurately reflect the various perspectives and interests of multiple stakeholder groups. Constituents (parents, students, educators, legislators, and employers of graduates) probably share some commonality with regards to what they expect from higher education, but they will probably also have unique viewpoints and issues they would want to have represented. An adequate indicator system should incorporate multiple perspectives through the use of multiple measures for each policy area and avoid aligning with any particular constituency group.

Appropriate standards of comparison address the need to establish clear benchmarks of progress or success based on peer performance, performance over time, established norms or best practices (Ewell & Jones, 1994). These standards of comparison should be specific to

institutional type (community college, liberal arts college, or research institution). For example, using an indicator of graduation may be adequate when compared against a benchmark from similar institutions or systems. Judging community college graduation rates by the same standard as four-year institutions, however, may not identify correctly a deficiency. When vocational students enroll in a very marketable program of study they may not graduate. They may complete only a couple of courses and already be in demand with employers. In this example, low graduation rates may not identify a deficiency but instead may identify programs that are filling an immediate need in the labor force.

Technical adequacy describes the need for indicators to be statistically valid and reliable, emphasizing their need to be statistically robust under typical policy conditions of missing or biased data (Ewell & Jones, 1994). Across systems of higher education, and even within individual institutions, missing or miscoded data are a common issue as are procedural changes in coding. Statistics employed in performance indicators must accommodate this or results will be questionable. The inclusion of educational researchers, those individuals with intimate knowledge about the data, could help to ensure technical adequacy.

Practicability refers to the requirement that indicators be attainable at a reasonable cost. This usually means that indicators must incorporate data that are already collected or can be collected without exorbitant expense. For outcomes assessments, one way to lower costs is to report on random, representative samples instead of the entire student population. Additionally, if results do not seem to significantly differ from year to year, assessments may not need to be annual and can, instead, be put on a rotating calendar.

Utilizing the aforementioned typology of indicators, including the criteria for judging indicator effectiveness, model indicator systems are subsequently identified and described.

Model Performance Indicator Systems

Tennessee was the first state to implement a performance indicator system. It is because of this long history of experience and willingness to try a wide range of indicators and budgeting and funding formulas that it has served as a model for more recent indicator systems across the nation. In 1976, Tennessee initiated performance funding with non-appropriated incentive funds that were allocated based on institutional performance on five indicators. At that time, the incentive pool was set at 2% of the total instructional budget. In 1984, the indicator list grew to nineteen and the pool was increased to 5%. In 1990, a new strategic plan established six areas of focus, with indicators for each. The following year, the indicators were reworked and became more quantitative and outcomes-oriented (Ewell, 1994). In 2001, Tennessee again reviewed its system and conducted a policy review of other states to identify model performance indicator systems (State of Tennessee, 2001). This review identified the following states as states with indicator systems that were worth emulating in some capacity: California, Connecticut, Florida, Georgia, Iowa, Illinois, Louisiana, Maryland, Mississippi, New Jersey, Oregon, South Carolina, South Dakota, Texas, and Washington.

Ruppert, (2004), concluded, "ten states have been among the leaders in developing new accountability policies and performance indicators: Colorado, Florida, Illinois, Kentucky, New York, South Carolina, Tennessee, Texas, Virginia, and Wisconsin" (p.1). A comparative analysis of these states, with respect to the manner in which each accountability process originated, the development of performance indicators, and for what purposes the results have been used, yielded some general conclusions about how indicator systems are linked to planning and budgeting. Tennessee's system has been driven by accountability, but

little consideration seems to have been given to the ability of indicators to provide insight about institutional improvement. The same seems true in Colorado, Florida, Kentucky, and South Carolina. These systems were developed in somewhat of a rush, with little guidance from researchers or statisticians that had worked with the data long enough to have an intimate understanding of what could or should be collected and reported. At the other end of the spectrum, extensive planning with input from multiple groups has been the focus in Illinois, New York, Virginia, and Wisconsin. These states focused on institutional improvement, not accountability. "At the time the case studies were undertaken, however, none of these states had been successful in producing the type of report sought by state policymakers, one that would allow for institution-level examination and comparison" (Richardson, 1994 p.134). The case study comparisons indicate that Texas has the best model for an indicator system in that it serves the dual purpose of public accountability and institutional improvement.

The Future of Performance Indicator Systems

In 2004, eleven states (California, Florida, North Carolina, North Dakota, Oklahoma, South Carolina, South Dakota, Tennessee, Texas, and Virginia) prepared statements for the National Commission on Accountability in Higher Education regarding lessons learned from and the future of performance accountability. Additionally, five principal researchers and policy analysts (Joseph Burke, Peter Ewell, Margaret Miller, Nancy Shulock, and Jane Wellman) were asked to prepare testimony. There are a number of general conclusions about what the future may hold that can be drawn from the expert opinions that were shared.

Conclusion 1: The future of performance accountability may see measures that are directly linked to state policy objectives. Fundamental to this, states would need to clearly assert their educational objectives and be willing to financially support their priorities. As Nancy Shulock (2004) described in her testimony to the Commission, "the mindset is still ultimately one of evaluating and comparing institutions with an eye to budgetary consequences, rather than examining state-wide educational performance for purposes of state policy intervention on a much broader scale" (p. 1). As she offered as an example, "institutions may all have good retention and graduation rates and yet the state may be educating far too few people to sustain a healthy civic and economic life" (p.2).

Conclusion 2: Performance accountability may, in the future, need to reflect issues of high priority at the state level. To focus efforts and send a signal of relative importance, systems should have a narrow list of indicators that are restricted to only those issues that are "amenable to government intervention" (Shulock, 2004, p. 2). Additionally, measures should avoid unnecessary duplication with other oversight initiatives in order to ease the institutional reporting burden. If measures are already reported elsewhere, state systems should obtain the information from those sources and preserve performance indicators for policy issues not already being addressed.

Conclusion 3: Future performance reporting systems should do what is right, not just what is easy. They should avoid using readily available data if it does not provide information that will be useful in decision-making (Burke 2004, Shulock 2004, and Wellman 2004). As previously mentioned, input indicators are easy to obtain but do not, by themselves, help to prioritize funding. Although it is not as easy as reporting readily available statistics, linking statistics back to policy is key to true accountability. For

example, reporting tuition and fees as an input measure only provides readily available data. Presenting tuition and fees as a percentage of state median income and then comparing the states would tie it to other input measures, other systems/states, and give it policy leverage (Burke & Minassians, 2002).

Conclusion 4: The future of performance accountability may see increased use of common measures. Up to this point, commonality has been limited to intrastate measures, if used at all. In 1995, the National Center for Education Statistics (NCES) pre-workshop inventory found that "institution-specific outcome measures did not translate well to such external audiences as state officials, parents, and students. Therefore, more consideration was being given to the development and reporting of common measures" (State of Tennessee, 2001, p.24). The future may focus on interstate commonality to make national comparisons possible. A state will then be able to judge its performance relative to other states.

Conclusion 5: Too much detail "overloads state accountability systems with far more data than people can digest and use by reporting data on individual institutional performance and mission-specific activities that are not helpful for state-level decision making" (Shulock, 2004). Future performance reporting should include concise conclusions with additional detail available for those who want it (Burke, 2004). These conclusions should include statements tying the indicator results back to policy and provide guidance for future priorities.

Conclusion 6: In the future, performance systems may focus on incentive funding (provided in advance of intended behaviors) as a more effective use of dwindling state resources than rewarding high performance (Shulock, 2004). This adjustment is a response

to institutions that perform at sub-standard levels but do not have the financial resources to implement change. Through this type of funding, institutions would have to show continuous progress toward a benchmark until it is achieved or risk loosing the funding.

Conclusion 7: The future of performance accountability may focus more on outcomes rather than inputs or processes and may more specifically include measurement of graduates' general education. In 2001, The National Forum on College-Level Learning recognized that most states were not reporting on the learning of their college graduates. This prompted a pilot study to implement a reporting model. The pilot was completed in Kentucky, and, subsequently, expanded to five more states. Without understanding college-level learning, "it is difficult to develop state-level policies and collaborations to serve underachieving subpopulations or regions of the state and to target resources accordingly" (http://collegelevellearning.org/projectsummary.htm). Nancy Shulock, from the Institute for Higher Education Leadership & Policy at California State University, emphasized in her testimony that the future of accountability should include "valid, aggregate measures of college-level learning." This belief was echoed in other testimony (Burke 2004, Wellman, 2004).

Measurement of College-Level Learning

The National Center for Public Policy and Higher Education's *Measuring Up 2002* report confirms there is a lack of national comparative data for college-level learning. "The degree to which students' knowledge and skills improve as a result of their education and training beyond high school is a key criterion for measuring state performance in higher education. All states receive an Incomplete in this category, as there is no information

available to make state-by-state comparisons" (measuringup.highereducation.org/2002/learning.cfm). The National Forum on College-Level Learning identified the importance of this lack of information for two primary reasons. First, without comparable data, there is no way to understand academic achievement on a national level. Second, these data are needed in order for policymakers in both the public and private sectors to monitor "educational-capital", which is defined as "the store of our nation's high-level knowledge and skills, which affects every contour of our political, economic, social, and cultural life" (Ewell, 2002, p. 2). Ewell stressed that understanding educational capital would help states prioritize funding, ensure equitable access to education, and build support for new initiatives. It would also allow policy makers to answer fundamental questions such as, "What are the competencies of students who graduate from postsecondary institutions?" and "How do each state's learning outcomes compare to national standards and to the outcomes achieved in other states?"

Measuring and Comparing Competencies

For community colleges, measuring competencies includes both vocational and academic outcomes. Vocational competencies are most commonly measured through participation and completion rates (overall and for non-traditional students), professional credentialing and licensing examination rates, employment placement and retention rates, average wage data, and satisfaction surveys (graduate, student, and employer). These data are reported to meet federal mandates of the Carl Perkins Applied Technology and Vocational Act and the Workforce Investment Act, as well as state performance indicator

systems and program review initiatives. Institutions, systems, and states could compare themselves against one another using these indicators of graduates' vocational gains.

The academic comparison of graduates is not as systematic. There are a variety of direct and indirect methods to measure academic outcomes. As Palomba and Banta (1999) described, "direct measures of learning require students to display their knowledge and skills" whereas indirect measures ask students to reflect on learning rather than demonstrate it (p. 11). Common indirect measures include surveys, focus groups, interviews, self-reported learning outcomes, and successful transfer. Direct measures include, but are not limited to, tests, essays, presentations, demonstrations, portfolios, and performance in capstone experiences. Arguably, the most common direct measures for institutional analysis are objective tests, either nationally or locally developed, because they are efficient to administer and score and have established validity and reliability.

There are advantages to both locally-designed and nationally-standardized tests. Locally-designed tests may reflect most accurately institutional goals but may be prohibitive due to the time and effort required to establish validity and reliability. In 1982, Florida locally designed and established use of the College Level Academic Skills Test (CLAST) as a rising junior exam. In 1988, it was for the "first time deliberately used in the aggregate to investigate questions of institutional effectiveness" (Banta, 2001, p. 342). In contrast to this locally-designed assessment, Tennessee and South Dakota adopted a nationally-standardized assessment of general education (ACT-COMP) for institutional accountability. Historically, institutions and states have opposed the use of nationally-standardized instruments on the grounds of institutional autonomy and individuality. This was evident in the early 1990s, as previously detailed, when states were beginning to mandate accountability reporting but still

allowed institutional discretion with respect to methodology. Institutional autonomy was (and in many cases is) valued above comparability. A decade later, the pendulum has swung the other way, and comparability across systems and states is now at the forefront of discussions (Miller, 2002, Ewell, 2002, National Center for Public Policy and Higher Education, 2002). Although, to date, there is no college-level learning outcomes assessment used systematically across states, Ewell (2002) reported there were six states (Arkansas, Florida, Georgia, South Dakota, Tennessee, and Texas) that have a common state-wide test for undergraduates. At that time, there were eight additional states (Colorado, Connecticut, Kentucky, Massachusetts, New Mexico, Rhode Island, Utah, and Virginia) that were in the process of developing a common approach to outcomes assessment.

There are numerous difficulties with implementing a state-wide assessment of college-level learning. As Ewell (2002) described, there are three political/organizational challenges to state-wide testing. First, maintaining state-wide consensus is difficult because of implementation costs and changing political priorities. Consensus is also a challenge due to the complexity (in policy and procedure) with implementing any state-wide initiative. The questions of who, what, where, and when must be resolved. The second organizational issue is that any common methodology must address the diversity of institutions, students, and expected outcomes across the state. As Ewell mentioned, higher education is unique from K-12 because there is little agreement about what outcomes ought to be expected and measured from graduates. This is especially true for general education, which incorporates some domains that are easily defined and measured and others that are not. The third organizational issue is that outcomes scores are a snapshot of students' performance at a given time and do not take into account the academic level at which students began their

postsecondary education. For these reasons, state-wide implementation of outcomes measurement is difficult. However, it appears the future of performance mandates will require this type of measurement ...and more (Miller, 2002, Ewell, 2002, National Center for Public Policy and Higher Education, 2002). The next step may be to embrace a common approach on a national level by either adopting an already existing instrument or by creating a valid, reliable measure of performance.

Assessments can measure relative performance, absolute performance, or gains in performance. There are instruments and techniques for each measurement goal, differing with respect to not only their purpose but also their content, item characteristics, and interpretation (Popham, 1975). Norm-referenced assessments compare students' relative performance, either by comparing the performance of individual students within the group being tested or by comparing their performance with a normative group representing others of similar age, experience, and background. The purpose of these assessments is to discriminate between high and low performance by distributing (ranking) students' average performance on a bell curve, with some students performing very well, most performing average, and a few performing poorly. The content focus is usually on breadth, as opposed to depth of knowledge. Items typically vary in difficulty and are included for their discriminatory ability and for their ability to increase test reliability. Scores for these assessments are usually expressed as percentiles, deciles, or grade equivalents. The most salient shortcoming with this type of assessment is that, if the only comparative standard is the average, there is no objective way of knowing whether that average reflects an accurate standard of what and how much students should actually be learning.

Criterion-referenced assessments are designed to address this shortcoming by measuring students' absolute performance against pre-determined criteria, without regard to their performance relative to one another. For these tests, students are informed of the performance standards and taught to succeed on related outcome measures. The content of these assessments is usually more specific than norm-referenced assessments. The items are chosen to obtain a sampling of students' skills, and scores are usually expressed as a percentage. Criterion-referenced tests help to eliminate competition and may, thereby, increase cooperation in teaching and learning across institutions, systems, and states. The major drawback with this type of testing is the potential for making comparisons to inappropriate criteria.

A gains approach, also referred to as a value-added or ipsative approach, compares each student's performance with his/her own earlier performance with the goal of determining whether any gains have been made. The primary advantage to this approach is that it generally provides a clearer picture than absolute or relative performance measures because it accounts for the student's entry level. A more detailed description of value/gains analysis is subsequently provided.

Defining Gains

As Benjamin and Hersh (2003) stated, determining excellence and quality in higher education should be measured, at least in part, by the degree to which institutions shape the abilities of students. This type of gains analysis is the next evolution of accountability for higher education (Ewell, 2002 and Shulock, 2004). The following review defines a gains analysis as it relates to general education at two-year institutions, with details about research

designs, psychometric issues, statistical approaches, and limitations. Additionally, examples of general education gains analyses in the community college sector are presented, as well as possible strategies for the future.

Ideally, gains describe what students' improved capabilities or knowledge as a consequence of their education at a particular college. This type of research may be classified as a single group pre-post quasi-experimental design (Gliner & Morgan, 2000). This type of analysis requires, at minimum, having an assessment of students' performance as they begin college and an assessment of those same students after they have had the full benefits of their education at the completion of their degree. The gain is then the difference between their performance at completion and their initial performance. Assessing the same student at two (or more) points in time allows each student to serve as his/her own control, virtually eliminating the effects of demographic variables such as socioeconomic status, gender, race/ethnicity, and intelligence. It is, therefore, expected from this type of design that any differences between scores can be attributed to the college experience or maturation and not variability among students. Described as such, gains analysis is not a new or different type of assessment. Rather, it is the use of one or more statistical techniques to mitigate, as much as possible, some extraneous variables while comparing initial student performance with performance at a later date.

In the community college sector, a gains analysis may be completed for either vocational or academic programs. It is most easily completed for programs or disciplines with very specific, measurable outcomes and is complicated by the heterogeneity of general education programs. For example, measuring the gains of an auto mechanic's certification or degree program is relatively basic. On the other hand, measuring the gains of a general

education program that cuts across the curriculum and includes concepts that are difficult to operationalize is much more challenging. These two applications of gains analysis are usually approached differently to accommodate their uniqueness. For example, while both assessments should be objective and standardized, vocational or discipline-specific assessments may be more applied, whereas general education designs usually incorporate norm-referenced paper and pencil tests. The focus of the current review will be on the general education application of a gains analysis. Specific to general education, a gains analysis attempts to answer the question of whether students leave the institution with more general education skills and knowledge than with which they arrived and, in the aggregate, may provide some indication to taxpayers about the efficacy of public higher education.

General Education Defined

"Historically, the focus of general education has been on providing broad exposure to the skills and attitudes that help graduates function in society, rather than on developing specialized knowledge about particular disciplines" (Palomba & Banta, 1999, p. 239). While most community colleges can probably agree on this very general definition of general education, there are as many specific definitions as there are institutions. Commonly, general education incorporates reading, writing, math, and science, as well as the general knowledge described by Seybert (2002) to include communication, ethics, problem solving, and cultural awareness. Some of these domains (reading, writing, math, and science) lend themselves more easily to definition and measurement than do others. This is one of the major complications in assessing general education: identifying and defining appropriate outcomes, even for the more enigmatic components.

Research Designs for a Gains Analysis

Ideally, the measurement of educational change in an individual would be conducted utilizing multi-wave (time-series) criterion based assessment. This is the design that K-12 reform has adopted (Drury & Doran, 2003). Higher education may, in the future, acquiesce to this preferred methodology; however, currently there are no nationally standardized criterion-referenced assessments for general education. The practical considerations for time series designs in adult student populations (student participation, cost, administration, etc.) are yet to be resolved. So, while this design may be ideal, most community colleges measuring gains (and these institutions are the exception, not the rule) utilize a two-wave (pre/post) design and a nationally standardized norm-referenced instrument. As Gliner and Morgan (2002) detail, the pre/post design is a poor quasi-experimental design for a variety of reasons, including the lack of randomization to treatment and control groups and the limited number of data collection points. However, it is still the most commonly accepted methodology currently in use in higher education gains analyses. For this reason, the following basic design options are specific to this approach.

In its most simplistic form, a gains analysis is the subtraction of an entrance score from an exit score with both scores representing identical scales. Hanson (1988) described this design as a "simple gains score approach" (p. 58). If this design utilizes an appropriately standardized objective instrument, the internal validity of the study is strengthened by the instrument's established content reliability and validity. Additionally, the primary advantage to this approach is that it employs an analysis technique that is simple to complete and to explain to an audience. The major disadvantage to this design is that students may exhibit a practice (also termed carryover) effect at the time of retest because

they have already taken the assessment. There are design solutions to accommodate the practice effect.

Utilizing parallel forms is one design adjustment that can be made to eliminate the practice effect. In this design, two different forms of the same test are administered to students. The difference in forms may come from varying the item order, altering the answer choices, or by using slightly different wording or scenarios in the questions. Using parallel forms requires establishing the relationship between the two forms. As Gliner and Morgan (2000) defined, a reliability coefficient of at least .80 between the two forms would be expected for this type of design. Additionally, the means from the parallel forms must be equated. As illustration of this design, Bohr et al. (1994) administered parallel forms of the Collegiate Assessment of Academic Proficiency (CAAP), developed by the American College Testing Program (ACT), at two points in time to assess the academic gains at a sample of community colleges versus those made at a sample of four-year institutions. Following the same methodology, Pascarella et al. (1995) included a larger sample to increase the generalizability of these findings.

Another design adjustment that can be made to accommodate the practice effect is to use similar tests with comparable scores. This can be accomplished through the use of two tests with the same scale or through the use of standardized scores (*z* scores) if the two tests are proven to assess the same constructs. Standardized scores, termed *z* scores, are statistical transformations that indicate how far and in what direction each score deviates from its distribution's mean. The *z* score is, therefore, expressed in units of its distribution's standard deviation and is considered a standardization transformation because the transformed distribution has a mean of zero and a standard deviation of one.

Psychometric Issue in Gains Analysis

One of the most serious issues for a gains analysis comes from the lack of appropriate instrumentation. As mentioned previously, ideal instrumentation would be nationally standardized, criterion-referenced, and designed to measure change as opposed to measuring static traits. As Hanson (1988) described "assessment instruments that measure static traits typically have high degrees of stability...[and] assume that the underlying structure of the construct being measured does not change with time" (p.57). This results in an inverse relationship between stability and an instrument's ability to indicate change; the higher the stability, the lower the reliability to indicate change. Furthermore, as an instrument's ability to detect gains increases, reliability of the instrument decreases. Hanson (1988) concluded, "it is difficult to have high change-score reliability without questioning whether a change in what is being measured has occurred" (p. 58). In addition to this psychometric issue, there is the statistical issue of error variance that must be reconciled in this type of research.

Statistical Approaches to a Gains Analysis

To understand the statistical approaches for a pre/post gains analysis (the most common design in higher education) and the debates therein, it is important to note that an observed gain is the summation of true gains and error variance. As Willet (1994) described, "because of the uncompromising vicissitudes of nature, when a test or rating instrument is administered the obtained measurement combines a measure of the person's true capability and whatever random error happens to accompany measurement" (p. 672). Over the years, there have been at least four major statistical approaches advocated for adjusting out the error found in pre/post research designs. The internal validity, which Cook and Campbell (1979)

define as "the approximate validity with which we can infer that a relationship is causal" (p. 37), is threatened in a gains analysis design in educational research, as it is in all applied research. There are, however, statistical techniques that can be incorporated into the research design to strengthen the internal validity by accommodating error variance.

There is considerable debate about the ability of difference scores to indicate true gains rather than observed or measured gains. Willet (1994) reviewed the debate and concluded that difference scores do, indeed, provide an accurate indication of gains. He detailed four reasons that difference scores have been "falsely condemned" (p. 672). First, he explained, low difference score reliability does not necessarily indicate imprecise measurement; it could indicate instead that the entire group of students is changing at nearly the same rate. Second, critics point to the correlation between initial status and difference scores as an inherent obstacle in such an analysis. Willet argued that the correlation is not an obstacle; it is an "almost inevitable fact of life" (p. 673) and should, therefore, be part of the analysis. Third, there is debate about the reliability of difference scores because researchers studying similar student populations in similar settings and using the same instruments have found varying degrees of correlation between the difference score and initial status. Willet refuted this point of argument by acknowledging that, since educational change is continuous, the degree (and direction) of correlation will depend on when the initial testing is completed, and it is because these studies began at a variety of points in time that they yielded varying results. Fourth, the relationship between change and initial status has been argued to be inherently negative. "Hence, measuring two groups over time, with existing first-time differences measured, the group that scores lower automatically shows greater change when measured the second time" (Hanson, 1988, p. 58). Willet (1994) refuted this

point by referring the reader to Thorndike (1966), who showed a positive correlation between difference scores and initial status. Due to the debate over the appropriateness of the difference score, alternative statistical techniques have recently been applied to gains analyses. "So, even though the difference score is not the outcast that many critics have claimed, several modifications of it have been proposed in order to estimate true change" (Willet, 1994 p. 673).

One modification is the creation of residual gain scores by using regression to establish predictive gains and then subtracting the observed gains. The creation of this technique is an attempt to level the playing field at which students start. While this approach may account for students' pretest scores, it can create confusion in interpretation because the residual gain score is produced in a metric different than that of the pretest or posttest. Another disadvantage of this approach is it statistically eliminates any shared variance between variables, thereby, possibly overestimating the contribution of one variable and underestimating the effect of the education (Hanson, 1988). Other statistical approaches used for gains analysis include multiple correlation and regression. Additionally, causal analysis (path analysis) is also used because it helps to explain indirect and direct influences and allows the researcher to virtually eliminate the effect of student background variables (socio-economic status, race/ethnicity, gender, intelligence, etc.) by entering them into the model. There are two potential issues with the application of this type of structural modeling. First, it is difficult to create an unbiased model. Second, "most causal models assume a unidirectional causal relationship" (Hanson, p. 60). Increasingly, educational researchers are advocating time series data collection and growth models to combat the issues inherent in the

measurement of change that statistics have yet to overcome in a pre/post research design. The higher education accountability movement has yet to adopt the advice.

Limitations of a Gains Analysis

Regardless of the research design, there are limitations to a gains analysis. First, as with any applied research, there is no way to accommodate the effects of maturation. However, large samples and reference groups can help to mitigate the impact. Second, it is a statistical technique and, therefore, subject to a margin of error. Results should be reported with a confidence interval or categorized in accordance with distribution characteristics (normal or non-normal). Also, as Greene (2002) stated, "there is more random error in measuring gains in test scores than in measuring the level of test scores" (p. 1) because there is measurement error at each point in time. The research design and statistical techniques for this approach can be complicated. Third, gain scores do not differentiate between true learning and teaching to the test. Fourth, the results are not causal because there are a variety of factors besides the quality of education that cannot be entirely controlled that may influence educational gains, including the maturation of students over time. With these points made, gains analysis is still a valuable compliment to traditional analyses by mitigating many extraneous variables known to influence outcomes data, such as socio-economic status, gender, and academic preparedness because each student, statistically, acts as his/her own control in the research design. For this reason, gains analysis is a "crucial tool in the accountability toolbox – despite its flaws" (Greene, 2002, p.1).

Use of Gains Analysis in the Community College

An internet review of the states seems to indicate that, currently, very few community colleges have incorporated a general education gains analysis into their assessment plans. Many of those that have a gains analysis utilize locally-designed instruments. While these approaches are highly institutionally specific, they lack the comparability of nationally-standardized tests. One example of this is the General Academic Learning Experience (GALE) assessment that was piloted in ten southern California colleges (Cohen, Schuetz, Chang, & Plecha, 2003). Another example is Florida's use of the College Level Academic Skills Tests (CLAST) previously mentioned (Van de Water, 1994). While possibly effective at the local level, these types of assessments lack national comparability. As with nationally-developed instruments, whether or not these assessments measure actual or relative gains depends on if they are criterion- or norm-referenced.

A nationally-standardized approach some community colleges are taking is to use the College Board's Academic Profile as a pre- and post-test for students' general education abilities. This methodology is subject to the practice effect mentioned earlier if both forms are identical. However, it is arguable that the practice effect is minimal if a student takes the entire two years to obtain his/her degree. Another approach a handful of two-year institutions have started using was developed at the American College Testing Program (ACT). ACT developed a methodology by which the ACT Assessment, a standardized, objective instrument designed to assess high school students' educational development and their ability to complete college-level work, can be linked to the Collegiate Assessment of Academic Proficiency (CAAP), a standardized college-level general education assessment. Though content specification of the ACT assessment and the CAAP test are similar, the tests

are dissimilar enough in terms of content and difficulty that they cannot be directly equated. Therefore, a simple gains score approach would not be valid. Instead, a linkage relationship is used, wherein gain levels are based on the student's CAAP decile conditioned on his/her ACT decile. Residual gain scores are then categorized as below expected, expected, or above expected. The drawback with both of these approaches is the use of norm-referenced tests instead of criterion-based instruments, which allows only for the identification of relative gains.

Future of Gains Analysis in the Community College

The future of gains analysis in the community college sector may include two strategies to strengthen the methodology and, therefore, the validity of the results and conclusions drawn from them. First, higher education may follow the advice of psychometricians, as the K-12 reform has, and multiple designate points in time for data collection. This would help by allowing the use of growth models in the analysis, thereby yielding a more accurate picture of observed gains. One way of implementing this is to utilize the end point of K-12 accountability as the beginning point for postsecondary accountability, and then to test students at two additional points in time: at the end of the sophomore year at a four-year institution or graduation from a two-year institution and again at graduation from a four-year institution. The future may also include instrumentation that would allow for state-to-state comparisons. This point was repeatedly mentioned in testimony to the National Commission on Accountability in Higher Education (Burke, Shulock, & Wellman, 2004). Furthermore, the future of gains analysis may include the development of new instrumentation that is criterion-referenced to yield absolute (as opposed

to relative) gains and specifically intended for use in change studies. Non-static instrumentation would enhance the validity of these kinds of accountability efforts (Hanson, 1988, p. 57).

The remainder of this review will narrow the focus of accountability, performance reporting, and gains analysis to the Colorado Community College System, specifically the system's largest institution, Front Range Community College.

The Colorado Community College System (CCCS)

In Colorado, the community college system is the state's largest and fastest growing system of higher education, serving nearly 120,000 students annually (http://www.cccs.edu). The System has governance responsibility for 13 state community colleges, assists in authority over four area vocational schools, two local district community colleges, and career and technical programs in over 150 school districts throughout the state. The organizational structure is typical of higher education systems in that the President is responsible to the Board (established in 1967 by legislative action) which, in turn, is responsible to the Colorado Commission on Higher Education (CCHE) (http://www.cccs.edu). The CCHE then reports directly to the Executive and Legislative branches of the state government. To provide a frame of reference, Figure 1 depicts the main campus of each of the 13 state community colleges.

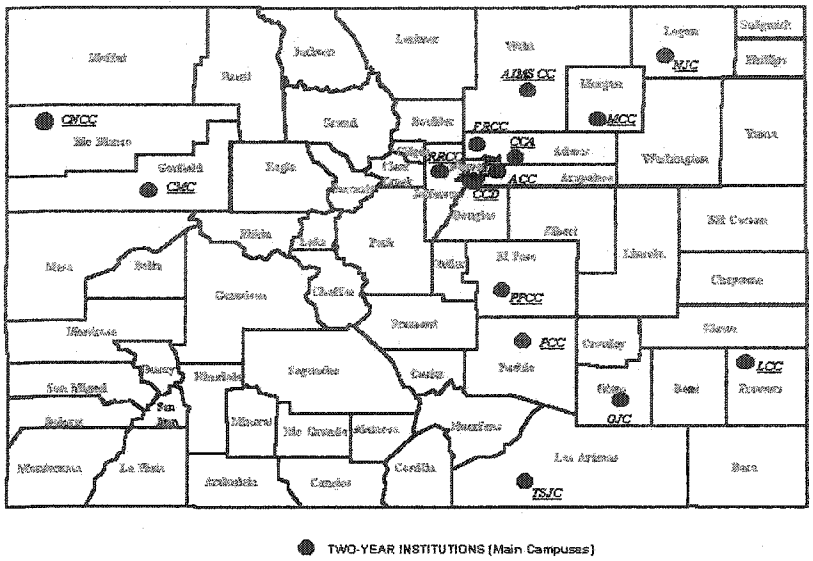

Figure 1. Colorado community college main campus locations.

Performance Indicators in the CCCS

The history of performance indicators in Colorado community colleges parallels the nation with respect to impetus, development, and future trends. Assessment came to Colorado at the same time it took the rest of the nation by storm. In 1985, the passage of Colorado House Bill 1187 required institutions to submit assessment and accountability plans to the Colorado Commission on Higher Education (CCHE). The legislation and the related CCHE policy mandated assessment but allowed the institutions methodological autonomy and discretion. The first data reported for this was for the 1989-1990 academic year. What resulted was similar to what happened in the rest of the country. Reams of institutionally-specific accountability reports were submitted and then sat on desks and went largely unread because of the amount of data and the lack of transparent conclusions.

When stakeholders had the data they requested from those accountability reports and still could not ascertain the relative performance of institutions, Colorado's first prescribed

performance indicator system was mandated. This occurred in 1996 with House Bill 1219, which directed the CCHE to develop a set of quality indicators. The resulting Quality Indicator System (QIS) was an attempt to mandate and prescribe what would be measured so institutions could be compared and the data submitted would be consistent across institutions and easier to understand. The QIS was modeled after many other performance indicator systems across the nation and prescribed ten indicators for community colleges as follows: three-year graduation rates (total population and minority), faculty workload, fall-to-fall retention (total population and minority), transfer rates, FTE expenditures, course availability, and credits to degree, as well as two institutionally specific indicators. In 1999, Senate Bill 229 expanded the QIS with additional indicators that were not tied to funding, as the original indicators were.

Amidst frustrations by colleges about the QIS, in 2003 the CCCS Institutional Research Advisory Group (IRAG) charged a subcommittee with the task of assessing the adequacy of the indicators for community colleges. The committee searched the literature to find best practices at peer institutions and similar systems across the nation. Several deficiencies in Colorado's current system were identified. First, there appeared to be a lack of appropriate comparison standards. Second, some of the indicators appeared to be excessively vulnerable to manipulation (intended or unintended). Third, not all of the indicators seemed appropriate at the community college level. For example, the faculty workload indicator seemed to be designed for research institutions to provide assurance that faculty were spending time teaching, not just researching. Fourth, it appeared the statistical methodology was flawed for some of the indicators. For example, employment rates were calculated including in the denominator those students who were not employed in their field

but who also were not attempting to find employment. This calculation may have led to erroneous conclusions about institutional effectiveness. Furthermore, it was discovered that, initially, colleges were not following standardized procedures for data reporting, serving only to exacerbate attempts at system-wide comparability.

Recommendations to correct these deficiencies (both in indicator appropriateness and operational definition) were made in a final report to the CCCS IRAG and the CCHE in the spring of 2003. Not only were colleges frustrated with the QIS at this time, other stakeholder groups were also frustrated with the amount of paperwork involved, the difficulty in creating institutional buy-in, and the inability of the system to provide transparent conclusions that could drive funding prioritizations at the state level. There did not seem to be a link between the QIS and policy or funding discussions, so the state is headed in a new direction.

The Future of Performance Reporting in the CCCS

In the closing days of the 2004 legislative session, Senate Bill 189, the College Opportunity Fund (COF) Act, was adopted. The Act is comprised of four interrelated components as follows.

First, the Act provides for stipends, which eligible students may use for undergraduate courses taken at a state or participating private institution. In-state students are eligible for stipends for up to 145 credit hours of undergraduate (college-level) course work. The exact dollar amount of the stipend will vary from year to year in response to the CCHE's recommendations to the legislature. The recommendations will be formulated based on inflation, enrollment growth, and faculty workload.

Second, the bill allows institutions to negotiate a fee-for-service contract to provide services for which students may not use stipend funds. For community colleges these services include Post Secondary Education Options (PSEO) credits.

Third, COF includes a provision to allow an institution (or group of institutions under a single governing board) to be given enterprise status if less than 10% of their funding is allocated by the state. Institutions with this designation would have increased flexibility in terms of tuition rates, fees, and bonding for capital construction.

Fourth, the legislation requires that governing boards negotiate a performance contract with the Department of Higher Education that will specify the performance goals the institutions will be required to achieve during the period delineated in the contract. The specific indicators of each contract are required to be measurable and reflect the role and mission of each governing board's institution(s). The CCHE Executive Director has explicitly stated three goals for the new performance indicator contracts. The indicators should focus on the knowledge and research outcomes an institution generates. In order to accomplish this, institutions should be able to measure academic (and vocational in the case of community colleges) gains. Additionally, the new indicators should be "transparent" (CCHE, 2004). They should be able to make a point to a diverse group of stakeholders, including but not limited to taxpayers, students, parents, and businesses, as well as communities the institutions serve. Furthermore, indicators should set high expectations. The idea behind performance contracts is to help raise the level of higher education quality in Colorado to a point where the institutions are national standards of excellence.

Summary

Colorado has, as has the rest of the nation, experienced many evolutions of accountability and performance reporting/funding in the recent past. The proposed study attempts to respond to the demands for general education accountability and demonstration of the academic gains made by college graduates. The study suggests a general education relative gains indicator that works within existing frameworks in Colorado's K-12 system and its higher education system to the largest extent possible in an effort to conserve resources and still obtain useful information to help guide future funding prioritization discussions for the community colleges.

Chapter III

METHODOLOGY

Research Questions

There are six research questions to be addressed with the proposed study.

1. Do FRCC graduates' ACT scores differ from the reference group?
2. Do FRCC graduates' CAAP scores differ from the reference group?
3. Do FRCC graduates' gains differ from the reference group?
4. Are FRCC graduates' gains equal across general education domain?
5. Are FRCC graduates' gains equal across campuses?
6. Are FRCC graduates' gains equal across degree granted?

Research Setting

The study was completed at Front Range Community College (FRCC) for three reasons. First, FRCC is a multi-campus, two-year, public, postsecondary institution serving both rural and suburban populations. It is, therefore, representative of the community college system. Second, with an annual unduplicated student headcount of 23,926 and over 9,948.76 student FTE, it is the largest community college in Colorado. Third, it is the only institution in the Colorado Community College System (CCCS) that utilizes any kind of standardized objective assessment of general education as a student completes a degree. The combination of the multiple locations, institutional size, and a unique assessment policy made it an ideal setting in which to pilot research to directly measure student learning in general education.

The institution is accredited by the Higher Learning Commission of the North Central Association of Colleges and Schools (FRCC, 2002).

Data Inclusion Criteria

The original data file provided by ACT included all FRCC students who took the CAAP in academic years 2000-01, 2001-02, 2002-03, and 2003-04 that could be matched back to their ACT assessment scores. The file (N = 978) included 848 graduates who were granted an AA, AS, AGS, or AAS degree. The remaining 130 students did not graduate with a degree, having either earned only a certificate, earned enough credits for a degree but never graduated, or had yet to complete their degree. Only the 848 graduates that were granted a degree were included in the study because only these students have the full benefit of the education provided at the institution.

Data Exclusion Criteria

Front Range Community College was able, until the 2004 academic year, to grant degrees through the Westminster or Larimer campuses only. Beginning in 2004, the college is also able to grant degrees through the new Boulder County campus. While all three campuses were included in the total data set, the Boulder County campus was excluded from the campus specific analysis due to small sample size (n = 7).

Data Management

All student demographic data are stored in the Student Information System (SIS), which is maintained at the Colorado Community College System (CCCS) office and

extracted, verified, and corrected as needed by FRCC staff at scheduled points in time through out the academic year. Questionable data are validated by admissions and records staff and corrected in SIS as needed. ACT assessment scores, CAAP scores, and the calculated gain levels were provided by ACT electronically and then matched against FRCC demographic extract files. The matching was based on the student's social security number and completed by the Director of Institutional Research at FRCC prior to being provided to the researcher.

Sample

The sample included 848 (302 male, 546 female) FRCC graduates from academic years 2000-01, 2001-02, 2002-03, and 2003-04. Of these, 96% ($n = 814$) were classified as in-state residents. The majority of the students in the sample were granted an A.A. degree (51.4%), 28.5% were granted an A.A.S., 13.1% were granted an A.S., and 7.0% were granted an A.G.S. degree. Of the 848 students, 11.5% self-reported at the time of application a racial/ethnic status other than Caucasian. The mean age at application was 22.0, with the median and mode at 22.0 and 21.0 respectively ($SD = 2.3$). There were six students over the age of 27, leaving 99.3% of the sample at age 27 or less.

Reference Group

For the current study, the reference group is defined as a national sample of public two-year postsecondary students who have taken both the ACT assessment and the CAAP test ($N = 23,999$).

Assessment Instruments and Gain Levels

Following are brief descriptions of the ACT Assessment, the Collegiate Assessment of Academic Proficiency (CAAP) test, and the corresponding domain-specific calculated gain levels. Each of the assessment instruments is a standardized objective measure of a variety of educational domains.

ACT Assessment: The ACT is a standardized, objective instrument designed to assess high school students' educational development and their ability to complete college-level work. The test covers four skill areas: English, mathematics, reading, and science (http://www.act.org/aap/). For the proposed study, ACT scores were provided directly from ACT.

Collegiate Assessment of Academic Proficiency (CAAP): The CAAP is a standardized general education assessment comprised of five objectively-scored domains: writing, mathematics, reading, critical thinking, and science. A writing essay test is also included in the CAAP. At FRCC, students are randomly assigned to either complete the objective form of the writing assessment or the essay form. For this reason, the sample size for the objective writing test is roughly half that of the other domains. For the proposed study, CAAP scores were provided directly from ACT.

Gain Levels:

> Though content specification of the ACT assessment and the CAAP test are similar, the tests are dissimilar enough in terms of content and difficulty that they cannot be equated. Instead, a linkage relationship will be used, wherein gain levels will not be based on scores but on a

student's CAAP decile conditioned on his/her ACT decile (Harmston, 2004).

Gain levels are domain specific and were provided directly from ACT. Only the CAAP domains of reading, writing, math, and science were included in a gains analysis because the critical thinking portion of the CAAP, as well as a writing essay portion, cannot be matched to the ACT. Calculated gain levels were categorized as below expected, expected, or above expected.

> Expected gains are defined as CAAP deciles falling in the middle 75% of the CAAP decile distribution conditioned on the ACT decile. A target of 75% coverage was used for several reasons. First, under normality assumptions, +/- one standard deviation accounts for 68% of the distribution. So the expected gain should be close to that 68% criterion. A second consideration was that other ACT programs used the 75% cutoff, and it was desired to have the ACT/CAAP definition of expected gain be consistent other programs. Finally, after statistically smoothing the ACT/CAAP bivariate distribution, 75% conditional intervals fit natural breaks in the data (Harmston, 2004).

Figure 2 provides sample data as a visual representation of this distribution for the reading domain with FRCC data. The additional three domains can be seen in Appendix B.

	CAAP Deciles										
ACT Deciles	1	2	3	4	5	6	7	8	9	10	Row Total
10	2	0	0	0	0	0	1	11	26	72	112
9	0	0	1	0	4	4	16	17	27	37	106
8	0	0	0	2	6	4	14	17	18	29	90
7	0	1	2	1	6	7	11	14	12	13	67
6	1	1	5	7	9	5	13	12	6	7	66
5	7	7	9	8	26	7	28	18	14	8	132
4	1	4	12	3	9	3	9	7	4	2	54
3	8	5	15	10	18	7	15	8	1	2	89
2	3	6	12	9	15	5	9	6	0	1	66
1	13	14	18	5	8	2	4	1	0	1	66
Column Total	35	38	74	45	101	44	120	111	108	172	848

Figure 2. ACT/CAAP matched distribution for the reading domain.

As labeled, the y axis indicates a student's ACT decile and the x axis represents a student's CAAP decile. Each cell indicates the frequency count for how many students scored at each ACT/CAAP decile match. In this example, 120 students had a CAAP decile of seven. Of these students, fifteen also had an ACT decile of four. In the figure, the shaded cells indicate expected gains, the area above indicates lower than expected gains, and the area below indicates higher than expected gains. The 15 students just mentioned exhibited higher than expected gains, with lower ACT deciles and higher CAAP deciles.

Statistical Methods for Each Research Question

RQ 1 and RQ 2. Eight one-sample *t* tests (one per domain per test) were used to test whether the sample's ACT and CAAP scores differ significantly from the reference group.

For the remainder of the study, the dependent variable (gain level) is ordinal (lower than expected, expected, and higher than expected). Stevens (1951) suggested that this limits the type of analysis techniques that can be legitimately employed because no relationship between data can be implied or assumed beyond that of rank-order. His view point describes one side of the measurement debate. Lord's (1953) analysis of football numbers showed

support for the other side of the debate, indicating that as long as data are normally distributed, then parametric techniques can be utilized with ordinal data. Gaito (1986) agreed with Lord and stated that proponents of Stevens' view

> apparently do not read the statistical journal literature, inasmuch as a number of articles on this topic showed clearly that measurement scales are not related to statistical techniques. Furthermore, they show little understanding of the mathematical statistics underlying the various statistical procedures. Scale properties do not enter into any of the mathematical requirements for the various statistical procedures (p. 564).

He proposed that the only assumption for parametric analysis that resembles scale is that of normality and, if data follow a normal distribution, irrespective of measurement, they would be of interval scale nature because the intervals between data points are known in terms of probabilities.

With both sides of the debate acknowledged, the more conservative, non-parametric techniques were utilized with the presentation of a parametric analysis using a continuous gain-level approximation score, which will be calculated by subtracting the ACT decile from the CAAP decile. Correlations for this computed approximate gain score with the ordinal gain levels are shown in Table 1.

Table 1

Correlations of Approximate Gains Scores and Gain Levels

Domain	*Correlation*	*p*
Reading	.760	.000
Writing	.762	.000
Math	.815	.000
Science	.715	.000

RQ 3. The one-sample chi-square test was used to test whether the sample gains differ from the reference group gains. After tabulating the observed counts at each gain level, the chi-square test procedure compared those counts with expected counts (calculated by applying the known reference group percentages to the sample). This non-parametric procedure was used because the gain levels are a categorical variable. A separate chi-square was used for each general education domain measured. A parametric analysis is not possible because the researcher does not have the ability to calculate an approximate gain score for the reference group subjects.

RQ 4. The non-parametric Friedman procedure was used to test whether gains are equal across general education domain. The Friedman was chosen because the samples are related and the dependent variable (gain level) is ordinal. Eight Mann-Whitney tests (one per domain) were utilized to compare genders and minority status. Within subjects ANOVAs were used with the calculated approximate gain score. An overall analysis was completed, as was comparison of genders and minority status.

RQ 5. The non-parametric Mann-Whitney test was used to test whether gain proportions are equivalent across campuses (Larimer, Westminster). This non-parametric procedure was chosen because the two samples are independent and the dependent variable (gain) is ordinal. Eight Mann-Whitney tests were utilized (one per domain) to compare genders and minority status. Between subjects t-tests were completed using the calculated approximate gain score. An overall analysis was completed, as was a comparison of gender and minority status.

RQ 6. The non-parametric Kruskal-Wallis test was used to test whether gain proportions are equivalent across degree granted (AA, AS, AGS, AAS). This non-parametric procedure was chosen because the four samples are independent and the dependent variable (gain) is ordinal. Eight Mann-Whitney tests were utilized (one per domain) to compare genders and minority status. A between subjects ANOVA was completed using the calculated approximate gain score. An overall analysis was completed, as was a comparison of gender and minority status.

Chapter IV

RESULTS

Sample

The data file provided by ACT included all FRCC students who took the CAAP in academic years 2000-01, 2001-02, 2002-03, and 2003-04 who could be matched back to their ACT assessment scores. The file ($N = 978$) included 848 graduates who were granted an AA, AS, AGS, or AAS degree. The remaining 130 students did not graduate with a degree and were, therefore, excluded from all analyses.

Descriptives

The sample included 848 (302 male, 546 female) FRCC graduates from academic years 2000-01, 2001-02, 2002-03, and 2003-04. Of these, 96% ($n = 814$) were classified as in-state residents. The majority of the students in the sample were granted an A.A. degree (51.4%), 28.5% were granted an A.A.S., 13.1% were granted an A.S., and 7.0% were granted an A.G.S. degree. Of the 848 students, 11.5% self-reported at the time of application a racial/ethnic status other than Caucasian. The mean age at application was 22.0 with the median and mode at 22.0 and 21.0 respectively ($SD = 2.3$).

RQ 1: Do FRCC graduates' ACT scores differ from the reference group?

One sample *t*-test results indicated FRCC graduates, across domains, had significantly higher ACT scores than the reference group; however, the effect sizes were small, indicating little practical significance. Figure 3 displays these data.

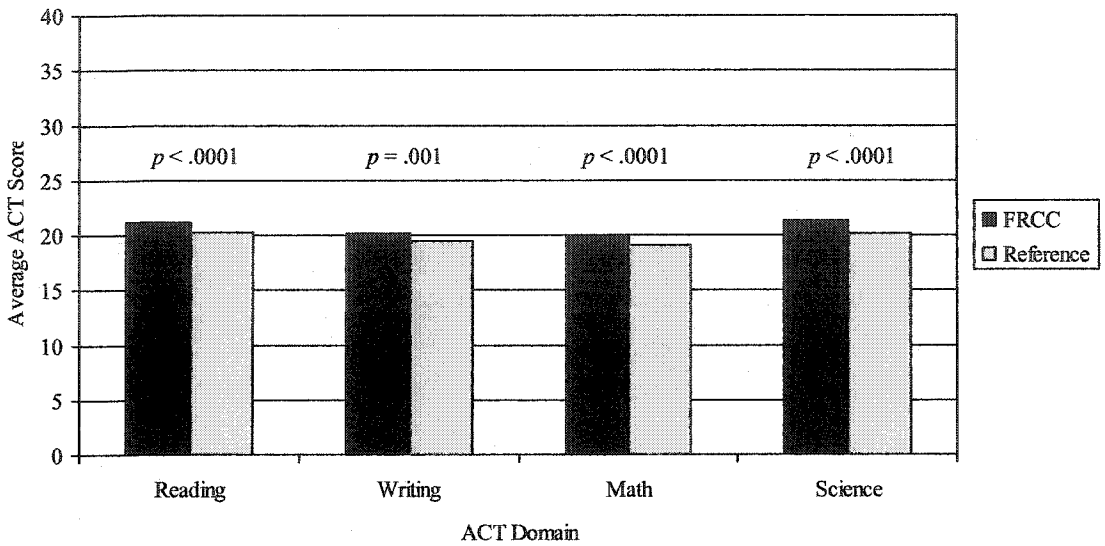

Figure 3. FRCC graduates' ACT scores compared with the reference group.

ACT Reading: FRCC graduates' mean reading score (21.12) was significantly higher than the reference group's mean score (20.3), with a mean difference of 0.817 points, $t(847) = 4.385$, $p < .0001$. Utilizing Cohen's (1988) interpretation of the *d* family of effect size, the effect size was small, $d = .15$. The statistically-significant result with a low effect size may be due to the large sample size.

ACT Writing: Graduates' mean writing score (20.19) was significantly higher than the reference group's mean (19.5), with a mean difference of 0.688 points, $t(552) = 3.431$, $p = .001$. The effect size was, again, small, $d = .15$.

ACT Math: Graduates' mean math score (20.01) was significantly higher than the reference group's mean (19.1), with a mean difference of 0.908 points, $t(847) = 6.862$, $p < .0001$. The effect size was small, $d = .23$.

ACT Science: Graduates' mean science score (21.25) was significantly higher than the reference group's mean (20.1), with a mean difference of 1.150 points, $t(847) = 8.863$, $p < .0001$. While the largest of the four domains, the effect size was small to moderate, $d = .30$.

RQ 2: Do FRCC graduates' CAAP scores differ from the reference group?

One sample *t*-test results indicated FRCC graduates scored significantly higher than the reference group on each domain of the CAAP. The effect sizes for the CAAP score results were larger than those for the ACT scores, suggesting, that, although FRCC graduates enter the institution at a similar point academically to their peers, they graduate at a somewhat higher point. Figure 4 displays these data.

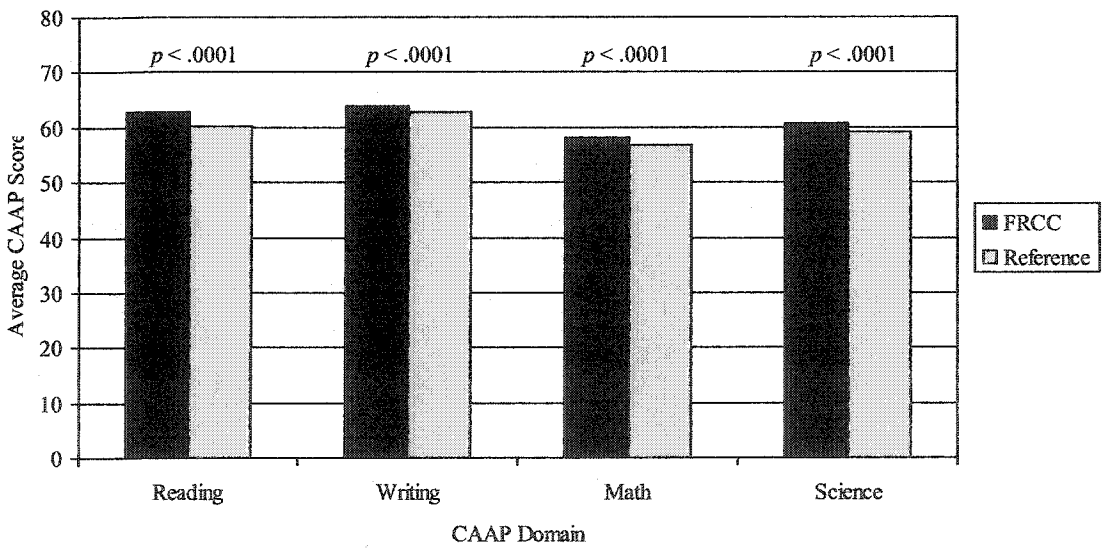

Figure 4. FRCC graduates' CAAP scores compared with the reference group.

CAAP Reading: FRCC graduates' mean reading score (62.73) was significantly higher than the reference group's mean score (60.3), with a mean difference of 2.433 points, $t(847) = 14.044$, $p < .0001$. The effect size was moderate, $d = .46$.

CAAP Writing: Graduates' mean writing score (63.9) was significantly higher than the reference group's mean (62.7), with a mean difference of 1.202 points, $t(552) = 6.383$, $p < .0001$. The effect size was small, $d = .26$.

CAAP Math: Graduates' mean math score (58.02) was significantly higher than the reference group's mean (56.8), with a mean difference of 1.218 points, $t(847) = 10.420$, $p < .0001$. The effect size was small to moderate, $d = .38$.

CAAP Science: Graduates' mean science score (60.61) was significantly higher than the reference group's mean (59.1), with a mean difference of 1.513 points, $t(847) = 10.342$, $p < .0001$. The effect size was small to moderate, $d = .36$.

RQ 3: Do FRCC graduates' gains differ from the reference group?

One-sample chi-square results indicated FRCC graduates' gains were significantly different from the reference group on each domain except writing, with alpha set at .0125 per domain (.05 for the entire research question).

Reading gains: Results indicated graduates' gains were significantly different from the expected proportions of the reference group, $\chi^2 (2, N = 848) = 125.140$, $p < .0001$. The proportion of graduates who had below expected reading gains ($P = .05$) was lower than the reference group ($P = .13$); the proportion of graduates who had expected reading gains ($P = .77$) was similar to the reference group ($P = .78$); the proportion of graduates who had higher than expected reading gains ($P = .18$) was higher than the reference group ($P = .09$). FRCC

graduates were 8% less likely to have lower than expected gains, 1% less likely to have expected gains, and 9% more likely to have higher than expected gains. Table 2 displays these results. Cramer's $V = .38$, indicating a moderate relationship.

Table 2

FRCC (Observed) Proportions of Reading Gain Levels Compared with the Reference Group (Expected)

	Proportions		
Gain Level	*Observed*	*Expected*	*Residual*
Lower than expected gains	*0.05*	*0.13*	*-0.08*
Expected gains	*0.77*	*0.78*	*-0.01*
Higher than expected gains	*0.18*	*0.09*	*0.09*

Writing gains: Results indicated graduates' gains were not significantly different from the expected proportions of the reference group, $\chi^2 (2, N = 553) = 7.325, p = .026$. The proportion of graduates who had below expected writing gains ($P = .07$) was lower than the reference group ($P = .09$); the proportion of graduates who had expected writing gains ($P = .81$) was similar to the reference group ($P = .82$); the proportion of graduates who had higher than expected writing gains ($P = .12$) was higher than the reference group ($P = .09$). FRCC graduates were 2% less likely to have below expected gains, 1% less likely to have expected

gains, and 3% more likely to have higher than expected gains. Table 3 displays these results. Cramer's $V = .12$, indicating a weak relationship.

Table 3

FRCC (Observed) Proportions of Writing Gain Levels Compared with the Reference Group (Expected)

	Proportions		
Gain Level	*Observed*	*Expected*	*Residual*
Lower than expected gains	*0.07*	*0.09*	*-0.02*
Expected gains	*0.81*	*0.82*	*-0.01*
Higher than expected gains	*0.12*	*0.09*	*0.03*

Math gains: Results indicated graduates' gains were significantly different from the expected proportions of the reference group, $\chi^2 (2, N = 848) = 45.732, p < .0001$. The proportion of graduates who had below expected math gains ($P = .10$) was lower than the reference group ($P = .13$); the proportion of graduates who had expected math gains ($P = .76$) was lower than the reference group ($P = .79$); the proportion of graduates who had higher than expected math gains ($P = .14$) was higher than the reference group ($P = .08$).

Table 4

FRCC (Observed) Proportions of Math Gain Levels Compared with the Reference Group (Expected)

	Proportions		
Gain Level	*Observed*	*Expected*	*Residual*
Lower than expected gains	*0.10*	*0.13*	*-0.03*
Expected gains	*0.76*	*0.82*	*-0.03*
Higher than expected gains	*0.14*	*0.08*	*0.06*

FRCC graduates are 3% less likely to have lower than expected gains, 3% less likely to have expected gains, and 6% more likely to have higher than expected gains. Table 4 displays these results. Cramer's $V = .23$, indicating a weak to moderate relationship.

Science gains: Results indicated graduates' gains were significantly different from the expected proportions of the reference group, $\chi^2 (2, N = 848) = 23.740, p < .0001$. The proportion of graduates who had below expected science gains ($P = .07$) was lower than the reference group ($P = .12$); the proportion of graduates who had expected science gains ($P = .82$) was higher than the reference group ($P = .80$); the proportion of graduates who had higher than expected science gains ($P = .11$) was higher than the reference group ($P = .08$). FRCC graduates were 5% less likely to have below expected gains, 2% more likely to have

expected gains, and 3% more likely to have higher than expected gains. Table 5 displays these results. Cramer's $V = .17$, indicating a weak relationship.

Table 5

FRCC (Observed) Proportions of Science Gain Levels Compared with the Reference Group (Expected)

	Proportions		
Gain Level	*Observed*	*Expected*	*Residual*
Lower than expected gains	*0.07*	*0.12*	*-0.05*
Expected gains	*0.82*	*0.80*	*0.02*
Higher than expected gains	*0.11*	*0.08*	*0.03*

RQ 4: Are FRCC graduates' gains equal across general education domain?

The non-parametric Friedman's results (using gain level) and parametric ANOVA results (using the computed continuous gain level approximation score) both indicated gains were significantly different across domains. There were no significant differences by gender or minority status.

Friedman's: Results indicated significant gain differences across domains, $\chi^2(3) = 17.501$, $p = .001$. Cramer's $V = .08$, indicating a weak relationship. Post hoc tests isolated the significant differences. Reading gains were higher than writing, math, and science (writing $z = -3.032$, $p = .002$, math $z = -4.195$, $p < .0001$, science $z = -5.120$, $p < .0001$). No other post hoc tests were significant.

Mann-Whitney (gender): Results indicated no significant differences across domain by gender (reading $p = .080$; writing $p = .611$, math $p = .523$, and science $p = .524$). Gender does not appear to differentially impact graduates' gains in any domain.

Mann-Whitney (minority status): Results indicated no significant differences across domain by minority status (reading $p = .929$; writing $p = .105$, math $p = .685$, and science $p = .065$). Minority status does not appear to differentially impact graduates' gains in any domain.

One-way ANOVA: Data violated both the assumption of normality (Kolmogorov-Smirnov $p = .000$ for each domain: reading statistic = 0.126, writing statistic = 0.432, math statistic = 0.102, and science statistic = 0.423) and the assumption of homogeneity of variance (Levene's $F = 525.70$, $p = < .0001$). However, ANOVA results indicated significant differences in gain across domains, $F = 204.89$, $p < .0001$. The effect size was moderate, $\eta = .41$. Table 6 displays the ANOVA results.

Table 6

Gain Differences by Domain

Source	*SS*	Df	*MS*	*F*	*p*	η
Between Groups	1,934.27	3	644.76	204.89	0.00	0.41
Within Groups	9,660.87	3,070	3.15			
Total	11,595.13	3,073				

Similar to the nonparametric post hoc tests, Bonferroni post hoc tests indicated reading gains were significantly higher than all other gains: writing $p = .002$, math $p = .015$, science $p = .004$.

Three-way ANOVA (gender, minority status, and domain): As displayed in Table 7, there were no significant main effects or interactions of gender or minority status with domain.

Table 7

Gain Differences by Domain, Gender and Minority Status

Source	*SS*	Df	*MS*	*F*	*p*	η
Corrected Model	1,953.81	15	130.25	4125	0.00	0.42
Intercept	1,627.34	1	1,627.34	515.40	0.00	0.39
Sex	0.01	1	0.01	0.00	0.97	0.00
Minority	0.10	1	0.10	0.03	0.86	0.00
Domain	581.52	3	193.84	61.39	0.00	0.24
Sex by Minority	0162	1	0.62	0.20	066	0.00
Domain by Sex	14.10	3	4.70	1.49	0.22	0.04
Domain bb Minority	3.93	3	1.31	0.41	0.74	0.00
Domain by Sex by Minority	1.76	3	0.59	0.19	0.91	0.00
Error	9,323.91	2,953	3.16`			
Total	15,580.00	2,959				
Corrected Total	11,277.72	2,968				

RQ 5: Are FRCC graduates' gains equal across campuses?

The non-parametric Mann-Whitney test results indicated gains were not significantly different across campuses, with one exception; math gains were significantly higher at the Larimer campus than the Westminster campus. The one-way ANOVA results were non-significant, indicating no differences by campus on any domain.

Mann-Whitney: As just stated, Mann-Whitney results indicated math gains were significantly higher at the Larimer campus than at the Westminster campus, $z = -3.223$, $p = .0013$. Cramer's V was .11, indicating a weak relationship.

Mann-Whitney (gender): Results indicated males at the Larimer campus had significantly higher gains than females at the Westminster campus in the math domain only, $z = -2.140$, $p = .032$. Cramer's V was .09, indicating a weak relationship.

Mann-Whitney (minority status): Results indicated no significant differences across campuses by minority status (reading $p = .090$, writing $p = .800$, math $p = .462$, and science $p = .882$). Minority status does not appear to differentially impact gains on either campus.

One-Way ANOVA: Data violated the assumption of normality (Kolmogorov-Smirnov $p = .000$ for each campus: Larimer campus statistic = 0.260, Westminster campus statistic = 0.250) but met the homogeneity of variance assumption (Levene's $F = 3.791$, $p = .052$). ANOVA results indicated gains were not significantly different across campuses, $F = 1.970$, $p = .161$.

Three-Way ANOVA: As displayed in Table 8, there were no significant main effects or interactions of gender or minority with campus.

Table 8

Gain Differences by Campus, Gender and Minority Status

Source	*SS*	Df	*MS*	*F*	*p*	η
Corrected Model	36.13	7	5.16	1.03	0.41	0.10
Intercept	167.76	1	167.87	33.55	0.00	0.20
Campus	0.04	1	0.04	0.01	0.93	0.00
Sex	8.77	1	8.77	1.75	0.19	0.05
Minority	0.02	1	0.02	0.00	0.96	0.00
Campus by Sex	4.27	1	4.27	0.85	0.36	0.03
Campus by Minority	2.63	1	2.63	0.53	0.47	0.03
Sex by Minority	2.04	1	2.04	0.41	0.52	0.02
Campus by Sex by Minority	0.01	1	0.01	0.00	0.97	0.00
Error	3,917.80	783	5.00			
Total	4,513.00	791				
Corrected Total	3,963.93	790				

RQ 6: Are FRCC graduates' gains equal across degree granted?

The non-parametric Kruskal-Wallis test results and the ANOVA results indicated significant gain differences across degree granted. The location of these differences is inconsistent across the two types of analysis.

Kruskal-Wallis: Non-parametric Kruskal-Wallis test results indicated math gains significantly differed across degree granted $\chi^2(3) = 51.4808, p < .0001$. Cramer's $V = .18$, indicating a weak relationship. There were no additional significant differences across degree granted. Post hoc pair wise comparisons indicated AS graduates had significantly higher math gains than all other degrees, (AAS $z = -6.489, p < .0001$; AGS $z = -2.863$,

$p = .004$; AA $z = -3.362$, $p = 0.001$). Additionally, AA graduates had significantly higher gains than AAS graduates ($z = -5.330$, $p < .0001$). There were no other significant pair wise comparisons.

Mann-Whitney (gender): Results indicated significant gender differences across degree granted for the math domain only, $\chi^2(7) = 53.365$, $p < 0.0001$. Cramer's V was .28, indicating a weak to moderate relationship. Regardless of gender, AS graduates had higher gains than AAS graduates. AA female graduates had higher gains that AAS graduates (male and female). AGS female graduates exhibited lower gains than AS graduates (male and female). AA male graduates had lower gains than AS females or AS graduates (male and female). Significant post hoc pair wise comparisons are presented in Table 9.

Table 9

Gain Differences in Math by Degree Granted and Gender

Relationship	z	p
AS Female > AAS Female	-5.228	<.0001
AS Female > AGS Female	-2.347	0.019
AS Female > AA Male	-2.933	0.003
AS Male > AAS Female	-4.761	<.0001
AS Male > AGS Female	-2.263	0.024
AS Male > AA Male	-2.643	0.008
AA Female > AAS Female	-4.993	<.0001
AA Female > AAS Male	-2.604	0.009

Mann-Whitney (minority status): Results indicated significant minority group differences across degree granted for the math domain, $\chi^2(7) = 57.706$, $p < .0001$. Cramer's

V =.20, indicating a weak relationship. Significant post hoc pair wise comparisons are presented in Table 10.

Table 10

Gain Differences in Math by Degree Granted and Minority Status

Domain	Relationship	*s*	*p*
Math	AA Non-minority > AGS Minority	-2.207	0.027
Math	AA Non-minority > AAS Minority	-2.126	0.033
Math	AA Non-minority > AAS Non-minority	-5.145	<.0001
Math	AS Non-minority > AGS Minority	-2.662	0.009
Math	AS Non-minority > AAS Minority	-6.131	<.0001
Math	AS Non-minority > AAS Non-minority	-3.64	<.0001
Math	AS Minority > AAS Non-minority	-2.08	0.038
Math	AS Minority > AAS Minority	-2.234	0.025
Math	AA Minority > AAS Non-minority	-3.052	0.002
Math	AA Minority > AAS Minority	-2.166	0.030
Math	AGS Non-minority > AAS Non-minority	-2.209	0.027
Math	AGS Non-minority > AS Non-minority	-2.552	0.011

One-Way ANOVA: Data violated both the assumption of normality (Kolmogorov-Smirnov $p = .000$ for each degree granted: AA statistic = 0.258, AGS statistic = 0.277, AS statistic = 0.212, and AAS statistic = 0.263) and the homogeneity of variance assumption (Levene's $F = 5.009$, $p = .002$). ANOVA results indicated significant gain differences by degree, $F(3) = 7.065$, $p < .0001$. The effect size was small, $\eta = 0.08$. Table 11 displays the ANOVA results.

Table 11

Gain Differences by Degree Granted

Source	*SS*	Df	*MS*	*F*	*p*	η
Between Groups	79.51	3	26.50	7.07	0.00	0.08
Within Groups	11,515.63	3,070	3.75			
Total	11,595.13	3,073				

The small effect size may indicate that the significant finding is an artifact of the large sample and the violations of the assumptions of ANOVA. Notwithstanding, Bonferroni post hoc tests indicated AS graduates had significantly higher gains than AA or AAS graduates ($p = .010, p < .000$, respectively) and that AA graduates' gains were higher than AAS graduates', $p = .021$.

Analysis by domain indicated no significant gain differences by degree for the writing domain. Significant differences were found in the other three domains. Within the math domain, $F = 20.143, p < .001$, the effect size was small to moderate, $\eta = 0.26$. Post hoc tests indicated AAS gains in math were significantly lower than all other degrees granted: AA $p = .000$, AGS $p = .012$, and AS $p = .000$. Within the reading domain, $F = 3.363, p = .011$, the effect size was small, $\eta = 0.11$. Post hoc tests indicated AGS graduates had higher reading gains than AAS, $p = .037$. Within the science domain, $F = 3.728, p = .011$, the effect size was similarly small, $\eta = 0.11$.

Three-Way ANOVA: As displayed in Table 12, there were no significant main effects or interactions of gender or minority status with domain. Neither gender nor minority status (alone or in combination) appears to impact gains on any domain.

Table 12

Gain Differences by Degree, Gender, and Minority Status

Source	*SS*	Df	*MS*	*F*	*p*	η
Corrected Model	100.56	15	6.70	1.77	0.03	0.09
Intercept	789.96	1	789.96	208.71	0.00	0.26
Degree	19.95	3	6.65	1.76	0.15	0.04
Sex	0.04	1	0.04	0.01	0.91	0.00
Minority	0.47	1	0.47	0.13	0.72	0.01
Degree by Sex	5.12	3	1.71	0.45	0.72	0.02
Degree by Minority	15.30	3	5.10	1.35	0.26	0.04
Sex by Minority	0.80	1	0.80	0.21	0.64	0.01
Degree by Sex by Minority	6.15	3	2.05	0.54	0.65	0.02
Error	11,177.16	2,953	3.79			
Total	15,580.00	2,969				
Corrected Total	11,277.72	2,968				

Summary of Results

RQ 1: Do FRCC graduates' ACT scores differ from the reference group?

FRCC graduates' ACT scores on each domain were significantly higher than the reference group. However, the practical significance of the findings was marginal, as exhibited by the small effect sizes. Results indicated that FRCC graduates enrolled in the institution with only slightly higher ACT scores than the reference group.

RQ 2: Do FRCC graduates' CAAP scores differ from the reference group?

FRCC graduates scored higher than the reference group on each domain of the CAAP. Graduates' scores on each domain were significantly higher than the reference group and displayed moderate effect sizes. Results indicated that FRCC graduates leave the institution with moderately higher CAAP scores than the reference group.

RQ 3: Do FRCC graduates' gains differ from the reference group?

FRCC graduates' gains differed from the reference group on every domain except writing. Compared to the reference group, a lower proportion of graduates demonstrated lower than expected gains across domains. Additionally, a higher proportion of graduates demonstrated higher than expected gains when compared to the reference group. The effect size for reading and science was moderate and was small for math. These results indicated, irrespective of ACT scores or CAAP scores, graduates demonstrated higher general education gains than the reference group on every domain except writing.

RQ 4: Are FRCC graduates' gains equal across general education domain?

Graduates' gains significantly differed by domain. Results indicated reading gains were significantly higher than gains in all other domains.

RQ 5: Are FRCC graduates' gains equal across campuses?

Nonparametric results indicated only one significant difference by campus; math gains were significantly higher at the Larimer campus than the Westminster campus. More

specifically, males at the Larimer campus had higher gains than females at the Westminster campus. Parametric results found no significant differences by campus.

RQ 6: Are FRCC graduates' gains equal across degree granted?

Non-parametric results indicated that AS graduates had significantly higher math gains than all other degrees and that AA graduates had significantly higher gains than AAS graduates. There were also some significant differences on the math domain by gender and minority status. The parametric results indicated no significant differences across degree within the writing domain but did indicate differences within the other domains.

Chapter V

SUMMARY AND CONCLUSIONS

An Existing Gap in Accountability

As previously chronicled, there have been many evolutions of assessment and accountability. What began in the 1970s and 1980s as the assessment movement evolved in the 1990s into mandated accountability. This shifted institutional motivation from voluntary internal improvement to mandated external reporting. The more recent evolutions of accountability to performance reporting, budgeting, and funding have retained this external motivation but have shifted the focus from inputs and processes to outputs and outcomes. While inputs refer to resources received to support the institution and processes refer to the manner in which services are delivered, outputs involve the *quantity* of what is produced, and outcomes refer to the *quality* of programs, activities, and services.

Despite these evolutions, every state received an "Incomplete" grade in The National Center for Public Policy and Higher Education's *Measuring Up 2002* report because of a glaring gap in research and practice in postsecondary accountability. As a nation, we are in desperate need of valid, standardized, aggregate measures of the general education with which students exit our institutions. Currently, without standardized data, there is no way to make interstate comparisons or understand academic achievement on a national level. These data are needed in order for policy makers in both the public and private sectors to monitor the educational-capital of our national economy (Ewell, 2002). Ewell stressed that understanding educational capital would help states prioritize funding, ensure equitable access to education, and build support for new initiatives. It would also allow policy makers

to answer fundamental questions regarding graduates' competencies and institutional effectiveness.

Although there has been much progress made in the area of postsecondary assessment and accountability, there is much left to do. In 2004, hearing testimony to the National Commission on Accountability in Higher Education focused on lessons learned from accountability, as well as the future of accountability in postsecondary education. Seven general conclusions were yielded from the testimony of five top educational researchers and representatives from eleven states. First, the future of performance accountability may see measures that are linked directly to state policy objectives. This direct link could provide substantive leverage for change in both the amount and process of funding to institutions or governing boards. Second, accountability may need to narrow its focus to only issues of high priority at the state level. Currently, many accountability systems are too broad and result in stacks of reports that may never get read. Third, accountability systems may require data beyond the easily calculated statistics of inputs and outputs. Data should be interwoven to create a full tapestry of a system's performance relative to state goals. Fourth, performance systems may incorporate incentive funding (provided in advance of intended behaviors) as a response to institutions that perform at sub-standard levels but do not have the financial resources to implement change. Fifth, accountability systems must be appropriate for state-level discussions. Currently, most systems include too much detail and are not as much help to policy analysts, legislators, and other higher education advocates as they could be. Sixth, the future of accountability may involve an increased use of common methodology and indicators across systems and states. This may allow for comparability across states. Seventh, the future of performance accountability may focus more on the measurement of

graduates' general education. The literature suggests that the future of performance accountability may include the use of standardized tests to provide the ability to make national comparisons across higher education institutions and systems. One researcher, in testimony to the National Commission on Accountability in Higher Education, identified use of the Collegiate Assessment of Academic Proficiency (CAAP) as one possible methodology for such comparisons. (Shulock, 2004, p.3).

An Attempt to Fill the Gap

The current study was intended to begin to fill the gap in the measurement of general education by utilizing the CAAP, in conjunction with the ACT, to measure relative gains. Use of the ACT and CAAP fits well into the existing general education and basic skills framework of Colorado K-12 and postsecondary education systems. This framework consists fundamentally of three pieces of legislation and their corresponding policies for implementation. First, all students enrolled in the eleventh grade are required to take the American College Test (ACT). Second, as part of basic skills assessment and mandatory placement at the postsecondary level, first-time, degree-seeking freshmen are asked to present their ACT assessment scores when applying for admission to any state supported community college. Third, the current higher education Quality Indicator System will be replaced in the next academic year with performance contracts which may, as mentioned at the Colorado Commission on Higher Education Accountability and Performance Summit in June 2004, include a measure of general education outcomes in the future. Utilizing the ACT and CAAP to assess these outcomes is fiscally efficient, eliminates duplicative testing for the

student, and provides national benchmarks for comparative purposes while anticipating the future of performance reporting in Colorado.

This study was conducted at Colorado's largest community college, Front Range Community College (FRCC), not only because of the college's size but also because of its multiple locations and unique assessment policy. FRCC is a multi-campus, two-year, public, postsecondary institution serving a variety of student populations. Furthermore, it is the only institution in the Colorado Community College System (CCCS) that utilizes any kind of standardized objective assessment of general education as students complete a degree. The combination of the multiple locations, institutional size, and unique assessment policy made it an ideal setting in which to pilot research to directly measure graduates' relative general education gains.

The research objectives for the study were to 1) close the feedback loop at Front Range Community College (FRCC) with a direct measure of general education relative gains in writing, reading, mathematics, and science skills and 2) to employ a nationally-validated methodology that can incorporate CCCS assessment policy to directly measure general education relative gains in writing, reading, mathematics, and science skills. The specific research questions were as follows:

1. Do FRCC graduates' ACT scores differ from the reference group?
2. Do FRCC graduates' CAAP scores differ from the reference group?
3. Do FRCC graduates' gains differ from the reference group?
4. Are FRCC graduates' gains equal across general education domain?
5. Are FRCC graduates' gains equal across campuses?
6. Are FRCC graduates' gains equal across degree granted?

Findings

Overall, the research objectives for the current study have been met. The use of the ACT and CAAP provided data and results that can close the assessment feedback loop for the college. The college can now answer questions about graduates' entry status, exit status, and infer general education gains relative to their peers across the nation. Additionally, the study has demonstrated the potential generalizability of the methodology to any institution within the CCCS through the incorporation of existing policies and processes. Furthermore, each research question has been addressed thoroughly with both non-parametric and parametric analysis techniques.

Summary of Results

The first research question asked if FRCC graduates' ACT scores presented at entry to the institution differed from their peers. It appears that FRCC graduates' ACT scores on each domain were significantly higher than the reference group statistically, but the practical significance of the findings was marginal as exhibited by small effect sizes. This is an important finding because it provides a frame of reference for the remainder of the study. FRCC graduates enter the institution only slightly ahead of the reference group. In the future, this type of analysis should be replicated with the entire student population at the college (not just graduates) to assist in course work design. Additionally, this type of information could help create a seamless K-16 accountability process. Furthermore, these results could help demonstrate the success of K-12 reform in the state if officials could show above average scores when compared to national benchmarks.

The second research question asked if graduates' CAAP scores differed from the reference group. FRCC graduates scored higher than the reference group on each domain of the CAAP and displayed moderate effect sizes. Results indicated FRCC graduates leave the institution with moderately higher levels of demonstrated general education knowledge and skills than the reference group. These results show favorably for the college and echo annual and trends analyses produced by the FRCC Office of Institutional Research that show that, for the last five years, graduates have been at or above national benchmarks on every domain of the CAAP. In the future, research should investigate the demonstrated performance of specific subpopulations as compared to national benchmarks and as compared to the general population at the college. With this type of research, the college could have a better understanding of what skills and knowledge graduates with various demographics are able demonstrate. This may influence future policy and practice with regard to student success.

The third research question asked if graduates' general education gains differed from the reference group. Results indicated graduates' gains differed significantly from the reference group on every domain except writing. Compared to the reference group, a lower proportion of graduates demonstrated lower than expected gains across domains. Additionally, a higher proportion of graduates demonstrated higher than expected gains when compared to the reference group. The effect size for reading and science was moderate and was small for math. These results indicated that, irrespective of ACT scores or CAAP scores, graduates demonstrate higher general education gains than the reference group on every domain except writing. These results, in particular, not only reflect favorably on FRCC, they also provide the policy leverage that has so far been lacking in the funding discussions at the state and system level in Colorado. In the future, this research could be

replicated at all Colorado community colleges to provide a systemic picture of the contributions of these institutions to the state's educational-capital. If this research were a system wide initiative, the Board could incorporate the measurement of general education into the new performance contracts, which currently (in draft form) include only outcome measurement of vocational skills and put off measurement of the general education component of the community college mission until some point in the future.

The fourth research question is much more specific to FRCC and asks if graduates' gains were equal across general education domains. Results indicated graduates' gains significantly differed by domain. Non-parametric and parametric results suggested reading gains were significantly higher than gains in all other domains. These findings are important for the institution to assess how general education skills are taught across the curriculum.

The fifth research question was also institutionally specific and asked if graduates' gains were equal across campuses. Nonparametric results indicated only one significant difference by campus; math gains were significantly higher at the Larimer campus than the Westminster campus. More specifically, males at the Larimer campus had higher gains than females at the Westminster campus. Parametric results found no significant differences by campus. These results are important because the institution needs to be able to ensure that, regardless of which campus a student attends, the education provided is of similar quality. Future research should include the newest degree-granting campus of the college, the Boulder County Campus.

The sixth, and final, research question asked if graduates' gains were equal across degree granted. This is an important question because the institution would like to see all graduates leave the institution with similar general education gains. Non-parametric results

indicated A.S. graduates had significantly higher math gains than all other degrees, and that A.A. graduates had significantly higher gains than A.A.S. graduates. There were also some significant differences on the math domain by gender and minority status. The parametric results indicated no significant differences across degree within the writing domain but did indicate differences within the other domains. These results are important because they can help to guide curriculum to ensure minimum basic standards for general education.

Limitations of the Current Study

The most salient limitation of the current study may be that it includes only FRCC graduates who can be matched back to an ACT score. This eliminates many students who do not come to the college with recent ACT scores. These students, if not exempted, take the Accuplacer. For these students who then have Accuplacer scores, not ACT scores, there is no way to link their entrance scores to their CAAP exit scores. For these students, it may be worth considering alternative methodologies. For example, the Accuplacer could be taken at entrance, at completion of basic skills course work (if any), and at graduation. The difference scores could then provide some indication of gains. This methodology, however, lacks national standardization and benchmarks but is a less expensive process than taking the CAAP.

Another limitation of the current study may be the reference group, comprised of students across the nation who took both the ACT and the CAAP. Colorado requires all eleventh graders to take the ACT, whether or not the student is planning on attending college. Colorado is one of only a couple of states with similar policies. This may leave the reference group heavily weighted by students who plan to attend college. The research group may

have more students in it that did not, at the time they took the ACT, intend to enroll in college. The reference group may, therefore, not be entirely representative of the research sample.

A further limitation of the current study may be the large sample size. The sample size was chosen because of the detailed gender and minority status analyses. However, utilizing a smaller sample size for the overall questions would reduce the power in the analyses and possibly eliminate significant statistical results with very small effect sizes. This could help to make the results more understandable.

An additional limitation may be the research design itself. As stated in the literature review, a time series analysis with a criterion-referenced instrument would be an ideal methodology. The current methodology measures a student only pre/post and utilizes norm-referenced instruments because it is an applied study and has real world constraints imposed on it such as instrument availability and cost. This may leave gaps in the measurement and allows for only relative gains analysis, not absolute.

Implications of the Current Study

Irrespective of the previously mentioned limitations, the study anticipated the future of accountability and provided pilot data with implications for future research and practice in the Colorado Community College System. The major implication is the use of general education relative gains as an institutional performance indicator.

This study utilized the ACT and CAAP linkage to measure these gains. This methodology has the benefit of national standardization, benchmarks, and basic analysis support for institutions or schools without staff to devote to such a project. Additionally, and

more specifically to Colorado, it fit well into the existing legislative and policy framework. If, in the future, all graduates from all Colorado community colleges were required to take the CAAP, just as all eleventh graders are now required to take the ACT, this methodology could be easily employed. The gains analyses could then be produced for each institution, as well as the system as a whole for presentation to key stakeholder groups. However, the cost of such an undertaking may be prohibitive. Alternatively, the same purposes could be served through the use of random samples from each institution. Results from this type of analysis embody the future of performance accountability and could have policy and, possibly, funding implications for the CCCS.

Furthermore, the study has implications on a broader level within the state. There are currently discussions taking place regarding the possibility of the seamless assessment of public education from kindergarten through the baccalaureate level. The methodology used in this study could serve as part of this seamless assessment by utilizing the K-12 end point as the postsecondary start point.

Additionally, the current study could have policy implications at the institutional level for FRCC. These results, and results from future research in this area of study, could include program evaluation, public relations, foundation support, and boosting morale of students, staff, and faculty. The same could be true for other institutions that would eventually incorporate the methodology into their assessment initiatives.

General Conclusions

The current study incorporates a nationally-validated methodology into an existing assessment framework to illustrate the general education abilities of graduates from

Colorado's largest community college. It yields institutionally-specific results and implications, as well as more generalizable implications for the future of postsecondary assessment in the state. Most importantly, it anticipates the future of postsecondary assessment on a national level and, if incorporated at the state level, could position Colorado to be a leader in the accountability movement.

BIBLIOGRAPHY

Abt, P. (2004, April). *Accreditation and assessment: The role of student services.* Pre-conference workshop presentation at the Front Range Community College Student Services In-Service, Longmont, CO.

Astin, A. (1987). Assessment, value-added, and educational excellence. In D. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain. New directions for higher education* (pp. 89-108). San Francisco: Jossey-Bass.

Banta, T. (1985). The use of outcomes information at the University of Tennessee, Knoxville. In P.T. Ewell (Ed.), *Assessing educational outcomes. New directions for evaluation* (pp. 19-32). San Francisco: Jossey-Bass.

Banta, T. (1999). *Assessment in community colleges: Setting the standard for higher education.* Boulder, CO: National Center for Higher Education Management Systems.

Banta, T., & Borden, V. (1994). Performance indicators for accountability and improvement. In V. Borden & T. Banta (Eds.), *Using performance indicators to guide strategic decision making. New directions for institutional research* (pp. 95-106). San Francisco: Jossey-Bass.

Benjamin, R., & Hersh, R. (2003). Measuring the difference college makes: The RAND/CAE value added assessment initiative. *Peer Review,* 5(4), 26-29.

Boatright, K. (1995). University of Wisconsin's system accountability. In G. Gaither (Ed.), *Assessing performance in an age of accountability: Case studies. New directions for higher education.* (pp.51-60). San Francisco: Jossey-Bass.

Bohr, L., Pascarella, E., Nora, A., Zusman, B., Jacobs, M., Desler, M., et al. (1994).

Cognitive effects of two-year and four-year institutions: A preliminary study. *Community College Review*, 22, 4-11.

Borden, V., & Bottrill, K. (1994). Performance indicators: History, definitions, and methods. In V. Borden & T. Banta (Eds.), *Using performance indicators to guide strategic decision making. New directions for institutional research* (pp. 5-22). San Francisco: Jossey-Bass.

Burke, J. (1998). Performance funding indicators: concerns, values, and models for state colleges and universities. In J. Burke & A. Serban (Eds.), *Performance funding for public higher education. New directions for institutional research.* (pp. 49-60). San Francisco: Jossey-Bass.

Burke, J. (1998). Performance funding: Present status and future prospects. In J. Burke & A. Serban (Eds.), *Performance funding for public higher education. New directions for institutional research.* (pp. 5-14). San Francisco: Jossey-Bass.

Burke, J. (1998). Performance funding: Arguments and answers. In J. Burke & A. Serban (Eds.), *Performance funding for public higher education. New directions for institutional research.* (pp. 85-90). San Francisco: Jossey-Bass.

Burke, J., & Minassians, H., (2002). *Reporting higher education results: Missing links in the performance chain. New directions for institutional research.* San Francisco: Jossey-Bass.

Burke J., & Serban A. (1998). State Synopses of Performance Funding Programs. In J. Burke & A. Serban (Eds.), *Performance funding for public higher education. New directions for institutional research.* (pp. 25-48). San Francisco: Jossey-Bass.

Burke, J. (2002). *Funding public colleges and universities for performance*. Albany, NY: Rockefeller Institute Press.

Burke, J. (2004, April). Testimony to the National Commission on Accountability in Higher Education.

Ciereszko, A. (1987). Mandated testing in Florida: A faculty perspective. In D. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain. New directions for higher education* (pp. 67-76). San Francisco: Jossey-Bass.

Cohen, A., & Brawer, F. (2003). *The American community college*. (4th ed.). San Francisco: Jossey-Bass.

Cohen, A., Scheutz, P., Chang, J., & Plecha, M. (2003). Assessing community college student knowledge in the liberal arts. *Journal of Applied Research in the Community College,* 11(1), 21-32.

Colorado Commission for Higher Education. (2004, May). Event invitation. *Proceedings of the Colorado Higher Education Summit on Accountability and Performance*

Cook, T. & Campbell, D. (1979). *Quasi-experimentation: Design and analysis issues for the field settings.* Boston: Houghton Mifflin.

Curry, W., & Hager, E. (1987). Assessing General Education: Trenton State College. In D. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain. New directions for higher education* (pp. 57-66). San Francisco: Jossey-Bass.

Drury D., & Doran, H. (2003, January). The value of value added. *National School Boards Association Policy Research Brief*, 3(1).

Earl, S., McConnell, M., & Middleton, I. (1997). *Postgraduate certificate in tertiary-level teaching: Module 1.* Retrieved November 12, 2003, from Robert Gordon University,

Center for the Enhancement of Learning and Teaching Web site: http://www2.rgu.ac.uk/subj/eds/pgcert/assessing/assess5.htm

Ewell, P. (1987). Establishing a campus-based assessment program. In D. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain. New directions for higher education* (pp. 9-24). San Francisco: Jossey-Bass.

Ewell, P. (1994). Tennessee. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.83-93). Denver: Education Commission of the States.

Ewell, P. (2002). *Grading student learning: You have to start somewhere.* Retrieved August 3, 2004, from measuringup.highereducation.org/2002/articles/peterewell.htm

Ewell, P. (1994). Wisconsin. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.117-129). Denver: Education Commission of the States.

Ewell, P. (1994). Virginia. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.147-166). Denver: Education Commission of the States.

Ewell, P. (2002, January). State-wide efforts to assess student learning outcomes: Observations. *State Higher Education Executive Officers Network News* 21(1). Retrieved July 27, 2004, from www.sheeo.org

Ewell, P. (1994). Developing state-wide performance indicators for higher education: Policy themes and variations. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.104-116). Denver: Education Commission of the States.

Ewell, P. & Jones D. (1994). Data, indicators, and the National Center for Education Management Systems. In V. Borden & T. Banta (Eds.), *Using performance indicators to guide strategic decision making. New directions for institutional research* (pp. 23-36). San Francisco: Jossey-Bass.

Ewell, P. & Jones D. (1994). Pointing the way: Indicators as policy tools in higher education. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.6-16). Denver: Education Commission of the States.

Ewell, P. (2002). Assessment in Colorado. In T. Banta (Ed.) *Assessment update: The first ten years.* (pp. 164-166).

Freeman, T. (1995). Performance indicators and assessment in the State University of New York System. In G. Gaither (Ed.), *Assessing performance in an age of accountability: Case studies. New directions for higher education.* (pp.25-50). San Francisco: Jossey-Bass.

Gaito, J. (1986). Some issues in the measurement-statistics controversy. *Canadian Psychology, 27*, 63-68.

Gaither, G., Nedwek, B., & Neal, J. (1994). *Measuring up: The promises and pitfalls of performance indicators in higher education.* (ASHE-ERIC Higher Education Report No. 5). Washington D.C: The George Washington University Graduate School of Education and Human Development.

Gliner, J., & Morgan, G. (2000). *Research methods in applied settings: An integrated approach to design and analysis.* Mahwah, N.J.: Lawrence Erlbaum Associates.

Greene, J. (2002). The Business Model. *Education Next,* 2, Retrieved October 20, 2003, from www.educationnext.org/20022/20.html

Halpern D. (1987). Student outcomes assessment: Introduction and overview. In D. Halpern (Ed.), *Student outcomes assessment: What institutions stand to gain. New directions for higher education* (pp. 5-8). San Francisco: Jossey-Bass.

Hanson G. (1988). Critical issues in the assessment of value added education. In J. Banta (Ed.), *Implementing outcomes assessment: Promise and Perils. New directions for institutional research.* (pp. 53-68). San Francisco: Jossey-Bass.

Harmston, M. (2004). (personal communication, June 19, 2004).

Lingenfelter, P. (2001, November). Educational accountability. *State Higher Education Executive Officers Network News* 20(3). Retrieved July 27, 2004, from www.sheeo.org

Lord, F. (1953). On the treatment of football numbers. *American Psychologist, 8*, 750-751.

Miller, P. (2002, January). Measuring up on college-level learning: Comments. *State Higher Education Executive Officers Network News* 21(1). Retrieved July 27, 2004, from www.sheeo.org

National Center for Education Statistics. (2002). *Digest of education statistics.* Retrieved November 14, 2003, from www.nces.ed.gov/programs/digest/d02/definitions.asp

National Center for Education Statistics (2004). *Remedial Education at Degree-Granting Postsecondary Institutions in Fall 2000.* Retrieved July 20, 2004, from nces.ed.gov/surveys/peqis/publications/2004010/index.asp

National Center for Education Statistics. (2004). *Condition of Education.* Retrieved August 2, 2004, from www.nces.ed.gov/programs/coe

National Center for Public Policy and Higher Education. *Measuring Up 2000: The State by-State Report Card for Higher Education.* San Jose, CA

National Center for Public Policy and Higher Education. (2002). *Measuring Up 2002.*

Retrieved August 3, 2004, from measuringup.highereducation.org/2002/learning.cfm

Nettles, M., Cole, J., & Sharp, S. (1997). *Assessment of teaching and learning in higher education and public accountability: State governing, coordinating board & regional accreditation association policies and practices.* Stanford, CA: Stanford University, National Center for Postsecondary Improvement.

North Central Commission on Accreditation and School Improvement. (2000). Retrieved January 5, 2004, from www.ncacasi.org/postsecondary

Palomba, C., & Banta, T. (1999). *Assessment essentials: Planning, implementing, and improving assessment in higher education.* San Francisco: Jossey-Bass.

Pascarella, E., Bohr, L., Nora, A., & Terenzini, P. (1995). Cognitive Effects of Two-Year and Four-Year Institutions: New Evidence. *Educational Evaluation and Policy Analysis*, 17, 83-96.

Popham, J. W. (1975). *Educational evaluation*. Englewood Cliffs, New Jersey: Prentice-Hall Inc.

Richardson, R. (1994). Illinois. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.40-51). Denver: Education Commission of the States.

Richardson, R. (1994). Texas. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.94-103). Denver: Education Commission of the States.

Richardson, R. (1994). The State University of New York. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.65-72). Denver: Education Commission of the States.

Richardson, R. (1994). Effectiveness in undergraduate education: An analysis of state quality indicators. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.131-146). Denver: Education Commission of the States.

Ruppert, S. (1994). Kentucky. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.52-64). Denver: Education Commission of the States.

Ruppert, S. (1994). South Carolina. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.73-82). Denver: Education Commission of the States.

Ruppert, S. (1995). Roots and realities of state-level performance indicator systems. In G. Gaither (Ed.), *Assessing performance in an age of accountability: Case studies. New directions for higher education.* (pp.11-25). San Francisco: Jossey-Bass.

Seybert, J. (2002). Assessment issues, methods, and policies. In T. Banta (Ed.) *Assessment update: The first ten years.* (pp. 249-250).

Seybert, J. (2002). Assessment of general education. In T. Banta (Ed.) *Assessment update: The first ten years.* (pp. 261-263).

Serban A. (1998). Precursors of performance funding. In J. Burke & A. Serban (Eds.), *Performance funding for public higher education. New directions for institutional research.* (pp.15-24). San Francisco: Jossey-Bass.

Shulock, N. (2004, May). Testimony to the National Commission on Accountability in Higher Education.

Sims, S. (1992). Student outcomes assessment: An historical review and guide to program development. Greenwood press: New York.

Smith, H., Armstrong, M., & Brown, S. (1999). *Benchmarking and Threshold Standards in Higher Education.* London: Kogan Page Ltd.

State of Tennessee (2001). *Measuring performance in higher education*. Nashville, TN: Detch, E., Wright, D., Hausser, G., Ogden, E., Hardin, C., Thurman, D. et al.

Statistical Package for the Social Sciences (Version 12.0) (2003) [Computer software]. Chicago: SPSS Inc.

Stein, R., & Fajen, A. (1995). Missouri's funding for results initiative. In G. Gaither (Ed.), *Assessing performance in an age of accountability: Case studies. New directions for higher education.* (pp.77-90). San Francisco: Jossey-Bass.

Stevens, S. (1951). Mathematics, measurement, and psychophysics. In S. S. Stevens (Ed.), *Handbook of Experimental Psychology*. New York: Wiley, pp. 21-30.

Van de Water, G. (1994). Colorado. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.7-26). Denver: Education Commission of the States.

Van de Water, G. (1994). Florida. In S. Ruppert (Ed.), *Charting higher education accountability: A sourcebook on state-level performance indicators.* (pp.27-39). Denver: Education Commission of the States.

Wellman, J. (2004, May). Testimony to the National Commission on Accountability in Higher Education.

Willet, J. (1994). Measurement of change. In T. Husen & T. Postlethwaite (Eds.), *International encyclopedia of education* (2nd edition) (671-678). Oxford: Elsevier Science Press.

APPENDIX A

RESPONSE TO REQUEST FOR USE OF FRCC EXTRACT DATA

Dear Laura,

This letter is in response to our discussion on August 24, 2004 regarding your use of Front Range Community College student data for analysis in your dissertation.

The research questions you will be investigating will benefit the college and the methodology ensures no identifiable student data will be either utilized or released. We are therefore able to provide you access to the data and look forward to the results.

Please work with the Director of Institutional Research to proceed.

Sincerely,

Janet Gullickson Ph.D.
President, Front Range Community College

Geri Anderson Ph.D.
Chief Academic Officer, Front Range Community College

Patricia Meade M.A.
Director of Institutional Research, Front Range Community College

APPENDIX B

ACT/CAAP MATCHED DISTRIBUTIONS

Reading

CAAP Deciles										
1	2	3	4	5	6	7	8	9	10	Row Total
2	0	0	0	0	0	1	11	26	72	112
0	0	1	0	4	4	16	17	27	37	106
0	0	0	2	6	4	14	17	18	29	90
0	1	2	1	6	7	11	14	12	13	67
1	1	5	7	9	5	13	12	6	7	66
7	7	9	8	26	7	28	18	14	8	132
1	4	12	3	9	3	9	7	4	2	54
8	5	15	10	18	7	15	8	1	2	89
3	6	12	9	15	5	9	6	0	1	66
13	14	18	5	8	2	4	1	0	1	66
35	38	74	45	101	44	120	111	108	172	848

Writing

CAAP Deciles										
1	2	3	4	5	6	7	8	9	10	Row Total
1	0	0	0	0	2	5	8	3	57	76
0	0	1	1	0	3	9	18	10	15	57
0	0	2	3	5	8	6	27	14	12	77
1	0	2	1	5	14	6	16	1	3	49
1	0	3	2	6	8	6	14	2	3	45
1	5	8	5	4	14	4	8	1	1	51
7	5	14	12	9	14	7	3	0	0	71
1	4	12	5	3	7	6	1	1	0	40
3	9	11	3	5	8	1	2	0	0	42
16	14	10	2	1	1	1	0	0	0	45
31	37	63	34	38	79	51	97	32	91	553

Math

CAAP Deciles										
1	2	3	4	5	6	7	8	9	10	Row Total
1	0	0	0	0	5	9	9	39	54	117
0	0	5	3	0	4	18	15	30	25	100
2	3	10	13	0	15	20	17	33	16	129
2	2	3	10	0	12	8	14	19	8	78
4	4	7	13	0	11	14	6	12	6	77
8	6	10	6	0	13	17	9	16	2	87
4	11	14	14	0	15	20	7	13	1	99
7	5	24	19	0	8	9	5	4	2	83
12	3	7	9	0	5	4	2	2	1	45
11	5	3	10	0	2	2	0	0	0	33
51	39	83	97	0	90	121	84	168	115	848

Science

CAAP Deciles										
1	2	3	4	5	6	7	8	9	10	Row Total
2	1	1	0	0	6	2	7	26	66	111
2	1	2	3	1	10	9	12	30	40	110
2	1	2	4	7	11	9	8	16	21	81
4	9	5	11	10	30	22	30	29	23	173
4	6	4	10	9	22	7	10	11	7	90
0	0	0	0	0	0	0	0	0	0	0
5	10	3	7	7	17	12	8	7	1	77
7	24	13	16	19	28	8	5	5	1	126
1	9	6	9	5	7	2	1	2	0	42
9	11	7	2	3	4	0	2	0	0	38
36	72	43	62	61	135	71	83	126	159	848