DEVELOPING AND USING A LOGIC MODEL FOR EVALUATION AND ASSESSMENT OF UNIVERSITY STUDENT AFFAIRS PROGRAMMING: A CASE STUDY

by

Jeff Cooper

Masters in Business Administration, Katz School of Business, University of Pittsburgh

Submitted to the Graduate Faculty of

the School of Education in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

SCHOOL OF EDUCATION

This dissertation was presented

by

Jeff Cooper

It was defended on

March 31, 2009

and approved by

Bill Bickel, PhD, Professor of Education

Dennis Galletta, PhD, Professor of Business, Katz School of Business

Maureen McClure, PhD, Professor of Education

Cindy Tananis, EdD, Professor of Education

Dissertation Advisor: John Weidman, PhD, Professor of Education

DEVELOPING AND USING A LOGIC MODEL FOR EVALUATION AND

ASSESSMENT OF UNIVERSITY STUDENT AFFAIRS PROGRAMMING:

A CASE STUDY

Jeff Cooper, MBA, PhD

University of Pittsburgh

This dissertation addresses theory and practice of evaluation and assessment in university student affairs, by applying logic modeling / program theory to a case study. I intend to add knowledge to ongoing dialogue among evaluation scholars and practitioners on student affairs program planning and improvement as integral considerations that serve mission and vision at the contemporary university. Insights on the following research questions can help determine theoretical justifications and forge an inventory of effective evaluation and assessment techniques in student affairs.

1.   How can logic modeling be used to analyze evaluations of student affairs programs and an overall assessment campaign?

2.   How might evaluators and planners have enlisted a logic model such as the one developed in this study to enhance the effectiveness of the assessment campaign at the profiled university student affairs unit?

These questions involve general principles and particular applications of my arguments in favor of using a logic model to analyze a comprehensive assessment campaign, as conducted by a designated student affairs assessment team. Although sets of workable techniques at one university may not generalize to another campus culture, findings will reveal how one institution

of higher education (IHE) has behaved and responded to new challenges and inputs – in this case, greater emphasis on evaluation and assessment to address issues of accountability and credibility for student affairs. Using logic modeling as the primary heuristic, this study analyzes what the university system depicted in case study has accomplished and might have accomplished. I also invite readers to join my speculation how using and perhaps customizing this logic model could guide the unit's next steps in ongoing assessment. If a logic model works retrospectively, then perhaps it might function proactively. My hope is that readers find descriptions and lessons to compare and contrast to their own evaluative practices, adding to the knowledge base and possible consensus about current practices for university student affairs assessment campaigns.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would be remiss in not mentioning perspectives gathered from a professional appointment in Residence Life, concurrent with my participation in the Social and Comparative Analysis program at the School of Education. While the job specifications might betray my natural quantitative bent towards analysis, the proximity to students lent towards a genuine affinity for education. Without this opportunity, a combination of networking and happenstance, I would perhaps not appreciate how students feel and how conscientious stewardship can help them each shape a worldview. This human element offers me impetus to continue along this career path and, ironically, cast aside some of the business oriented instincts that had originally led me to that academic program.

Besides, Residence Life paid most of my tuition....

Thanks :

To Peggy for being patient as humanly possible, and for sacrificing any and all of life's amenities and enjoyment for the privilege of being addressed as "Doctor & Doctor" on the next sets of numerous wedding invitations we shall receive; extra thanks and kudos for envisioning logic model configurations other than the convoluted flow charts in earlier iterations, giving rise to the fun blue pyramid.

xo j

CT – for fresh directions, assertive critique, and generosity with knowledge;

DG – of paramount concern in business, for adding great value to these proceedings;

JWii – for unyielding support ranging from funding to patient guidance. I have been

fortunate to have you as my program advisor and committee chair.

Thanks to each and all of you for this vote of confidence.

jcc

1.0   INTRODUCTION

This dissertation splices fresh footage into the ongoing picture show of current theory and practice of evaluation and assessment in university student affairs. By applying logic modeling – a visual representation of "program theory" – to a case study, I demonstrate the extent that a student affairs assessment team succeeded in evaluating a sequence of intended outcomes. I test how logic modeling presents a strategy that can enable planners and evaluators to organize and explain the various interlocking elements comprising a complete assessment campaign. Findings may deliver insights and perspectives to scholars and practitioners debating and facilitating the acceptance of assessment of student affairs as an integral consideration in mission and vision at the contemporary university.

1.1   PURPOSE OF STUDY

The following research questions encapsulate the purpose of this study:

1.   How can logic modeling be used to analyze evaluations of student affairs programs and the overall assessment campaign?

2.   How might evaluators and planners have enlisted a logic model such as the one developed in this study to enhance the effectiveness of the assessment campaign at the profiled university student affairs unit?

1

My research questions involve general theoretical justifications and particular applications of my arguments in favor of using a logic model to analyze and perhaps to structure a comprehensive assessment campaign in university student affairs. In critiquing how the subject of the case study performed in its assessment campaign, this dissertation addresses current and favored practices in evaluation and assessment in a university student affairs setting. Although sets of workable techniques at one university may not translate perfectly to another campus culture, responses to the research question may help forge a checklist on overall performance of evaluation and assessment teams in student affairs. That is, this study purports to accomplish meta-evaluation[1] of the case study and meta-analysis[2] of theory and practice in student affairs assessment.

1.1.1 **Objectives**

This study presents a basis for logic modeling that can be used to contextualize and meta-evaluate a university student affairs unit's assessment campaign. Although sets of workable techniques at one university may not generalize to another campus culture, findings will reveal how one modern institution of higher education (IHE) has behaved and responded to new challenges and inputs – in this case, greater emphasis on evaluation and assessment to address issues of accountability and credibility for student affairs. Using logic modeling as the primary heuristic, analysis weighs what the university system depicted in case study has accomplished. I describe how a logic model can be used to identify and describe procedural, contextual, and

---

[1] Meta-evaluation is an evaluation of an evaluation (Patton, 1990, p. 147), not synonymous with meta-analysis (Weiss, 1997).
[2] Weiss (1997) defines meta-analysis as aggregate results of evaluations of similar programs (p. 133). Later in sub-chapter 2.1.4, p. 25, I discuss this idea in more depth as a function of an evaluator's numerous possible roles.

stakeholder details unique to a particular university that uses evaluation and assessment to improve or reformulate programs.

I develop and apply an original logic model to an actual case study, both by retrofitting and by extrapolating. That is, I use the logic model to analyze and sometimes challenge actions taken by a university student affairs assessment team. My model borrows structural and mechanical features from W. K. Kellogg Foundation (2004, 2007), Asian Development Bank (ADB) (2007), and the University of Wisconsin-Extension (2008). I draw justification for logic models from Weiss (1997), Patton (1990), and other theorists.

The signature element of this university's campaign to create a "culture of assessment" in student affairs was determination of 14 Student Baseline Outcomes. (See Appendix A for the official list of Student Baseline Outcomes.) Actualizing this "culture of assessment" is the unit's proclaimed mission. The drivers of every intended outcome are this mission along with the vision "to provide [u]niversity… students with the best collegiate experience in the country".

Using a new logic model to organize and demonstrate knowledge flows within and between outcomes, my study 1) analyzes and evaluates the extent that this university Student Affairs unit has achieved its goals, and 2) suggests steps towards further success. An underlying premise is that following and perhaps customizing this logic model could guide next steps in ongoing assessment.

1.1.2 **Context**

This study involves two primary realms, assessment and student affairs. I confine the case study to the assessment process in the Student Affairs division at the University of Pittsburgh, a research institution in the United States, Middle Atlantic region. Description of the actual case

3

involves university and student affairs structures and decision-makers' rationale for formal assessment, in light of evaluation standards, accountability, ethics, and touching upon accreditation standards – from which until recently student affairs had been somewhat insulated. I present the case study from the viewpoint of a participant-observer, as co-chair of a formal Assessment & Measurement Team (hereafter, "the Team") from August 2005 to June 2007 – two complete seasons of an ongoing assessment cycle. I think that my immersion and actual stake in the profiled assessment campaign contributes richness beyond a usual, perhaps journalistic narrative.

Within Student Affairs departments[3], independent evaluations took place on a regular basis. While these projects were part of an overall picture of university department programmatic and academic assessment contributing to stated mission and vision, **this study focuses on the Team activity centered around Student Baseline Outcomes**. The diagram below illustrates how unit and team level evaluations fit together, using Residence Life (my professional base) as an example. ***The green box represents the scope of this study***; every other element in this diagram represents a related but separate topic, perhaps for future research.

---

[3] Student Affairs consists of Student Life, Residence Life, Career Services, Student Counseling, the International Office, Disability Services (DRS), Student Health Services, Judicial Services (USJS, absorbed into Student Life as of summer 2006), and Office of Cross-Cultural and Civic Leadership ("O3CL", established October 2006). Each Student Affairs unit is led by a Director who reports to the Dean.

Figure 1   **Project scope**
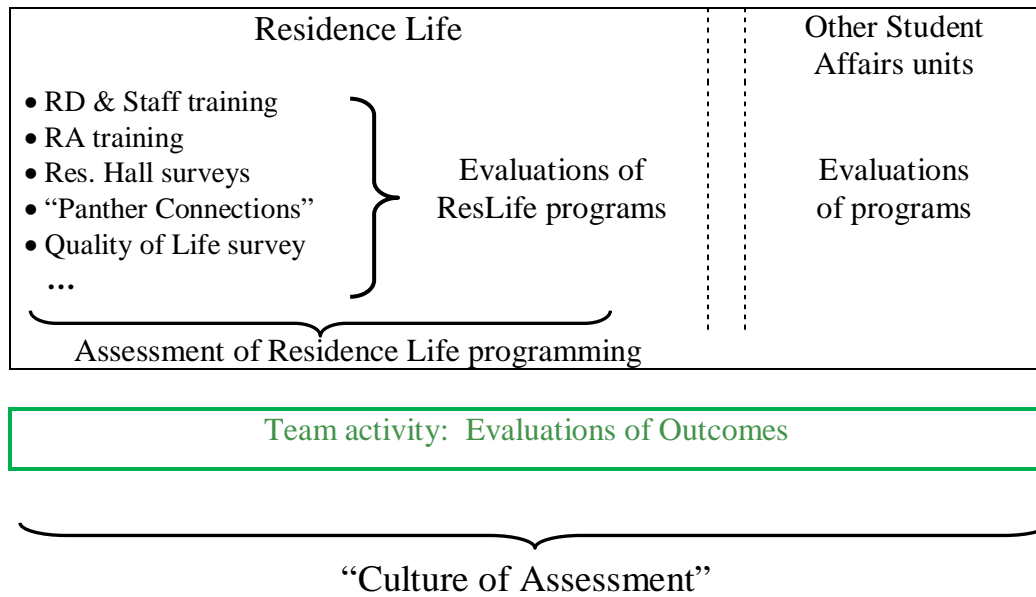
Chapter 4.1 will analyze facts of the case in greater detail. For now, the following timeline offers a preview and clarifies context.

Figure 2   **Timeline for Pitt Student Affairs assessment campaign, 2005-07**

Date / time frame	Event
July 2005	Dean of Student Affairs releases "Student Baseline Outcomes" compiled by task force.
August 2005	Dean designates five new teams to perform strategic initiatives in Student Affairs, including Assessment & Measurement Team.
September 8, 2005	Team conducts first meeting. The Team would meet the first Tuesday of every month until summer hiatus, and ad hoc for project preparations.
Oct. – Nov. 2005	Team conducts internal & external scans for e-portfolio and mentoring outcomes.
December 9, 2005	Team submits evaluation of e-portfolios.
December 23, 2005	Team chairs give festive and informative status update at annual Student Affairs Holiday Event.
January 31, 2006	Team submits evaluation of mentoring programs.
February 9, 2006	Team submits bundled evaluation report on internships, Pitt Pathway, Pitt Promise.
February 21, 2006	Team submits evaluation of parents conduit (later revised).
March 22, 2006	Team submits report on interviews with students to satisfy "listening" outcome.
May 18, 2006	Team follows up with report on Focus Groups based upon interview findings. Team submits report on Student Affairs "performance" Outcomes.
May 19, 2006	Team submits revised evaluation of parents conduit.
May 25, 2006	Chairs submit to Dean summary recommendations on programs.
June 1, 2006	Residence Life submits annual Quality of Life report, a source of data and quantitative support used by the Team and other Student Affairs planners.
June – August 2006	Summer hiatus.
September 2006	Chairs and liaisons agree to delay Team meetings until instructions from Dean.
October 24, 2006	Team meets for renewed campaign and plans evaluation activity, focusing on pending outcomes.
February 6, 2007	Team submits evaluation of Faculty resources and communications.
March, April 2007	Team deliberates, conducts scans on Conduct and Alcohol Violations.
June 5, 2007	Residence Life submits annual Quality of Life report to Dean (my final act in these roles).

1.1.3 **Plan**

This sub-chapter summarizes how I intend to 1) explain why these topics deserve attention, 2) describe in detail the Team's activity and contributions to the Student Affairs assessment campaign, and 3) present an original logic model as a framework to meta-evaluate the Team's performance, in response to the study's research questions.

The "Review of the Literature" chapter presents brief arguments affirming the strategic merits of assessment and of student affairs. Appraisal of student affairs focuses on historical origins, its justification under student development theory, emphasis on programming, and its gradual acceptance as an integral part of campus operations and activity. A premise is that student affairs is indeed worthy of the rigors of evaluation and assessment (Upcraft & Schuh, 1996; Schuh & Upcraft, 2000). Rather than another in-depth discussion of the histories of assessment and evaluation, of student affairs, and of these subjects' intersection, the "Review of the Literature" chapter addresses topics and issues that I believe planners and evaluators should consider when they enlist or construct a logic model to analyze and possibly guide an assessment process in student affairs. These topics include the political environment of assessment, buy-in among faculty and other stakeholders, assessment standards, accreditation issues, and the evaluator's role in building consensus and mutual trust. I conclude the literature review with origins of program theory and its embodiment as logic modeling, with Weiss (1997), Patton (1990), House and Howe (1999), latest updates from the Kellogg Foundation (2007), and a recent book by Frechtling (2007) as sources of wisdom and precedent.

The "Methodology" section unites these topics. Using Stake's (1978, 1994) guidelines, I employ narrative case study to illustrate how evaluators might measure program success and avoid pitfalls. Case description focuses on how a formal Assessment Team conducted

methodical program evaluation of "student baseline outcomes"[4]. Primary sources include communications memoranda, meeting minutes and agendas, formal reports and declassified documents, and descriptions of data gathering processes – i.e., interviews, focus groups, surveys, internal and external scans.

An important point is that the Team did not know about or use logic modeling in structuring or organizing activities to evaluate programs and measure achievement of baseline outcomes. Thus the "Methodology" section maps out how I retrofit case components and attributes into the logic model – a process that will allow me to critique Team accomplishments and activity and perhaps to offer insights how the Team could move forward. My hypothesis is that a logic model could have helped our Team make evaluations more organized and productive.

Enlisting logic modeling as the primary heuristic, my analysis adds knowledge to the field of evaluation by demonstrating how program logic enables planners and evaluators[5] to 1) use indicators and measurements to link intended outcomes from inputs; 2) frame assumptions within environmental influences, including peer and aspirant comparisons; and 3) categorize elements and processes into integrated, comprehensive checklists. The following figure shows the logic model I have designed to accomplish these tasks.

---

[4] These "baseline" outcomes are separate and independent from "learning outcomes" tracked in Academic Affairs, although planners envisioned both sets of outcomes to be complementary.
[5] "A term embodying all the charisma of science" (Chelimsky, 1997c, p. 101).

Figure 3   **Logic model depiction of assessment cycle**
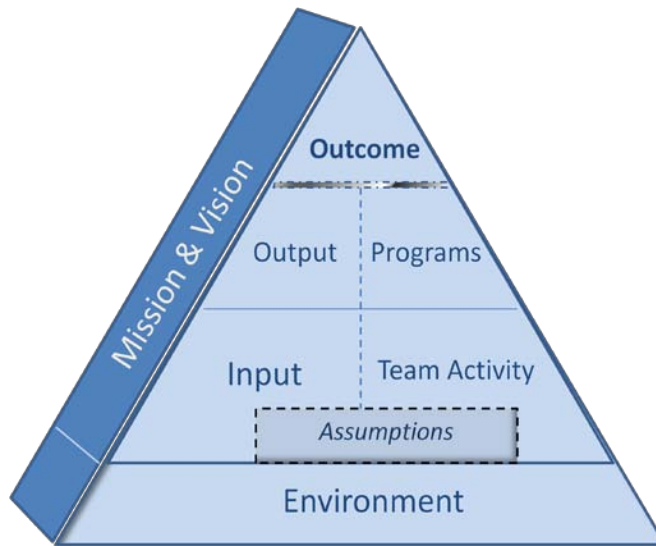
As the "spine" of the assessment process, mission and vision drive activity and frame outcomes, within environmental constraints. Underlying assumptions about inputs, outputs, and favorable courses of interrelated activities determine program formulation and implementation. The model offers a high-level view of process and linkages flowing upward, culminating in achievement or status of short-, medium-, and long-range outcomes. The shaded area between outputs and outcomes represents causality, which must be proven experimentally, presumed, or left to speculation lest outcomes be rendered "unintended" (Patton, 1990; Weiss, 1997; House & Howe, 1999; Upcraft & Schuh, 1996). Impact is represented by a trickle-down effect of outcomes back upon reconfigured assumptions and follow-up evaluations.

A wealth of documents and institutional knowledge provide facts, activities, and process flows that "fit" into the figure above via corresponding lists and detailed narrative. Such narrative comprises this study's "Analysis" chapter. The Student Affairs office depicted in this case study defined 14 outcomes. Hence, complete analysis could consist of 14 sub-chapters. Brevity and economy instead call for a representative sample of outcomes: I distill analysis into a

manageable three sub-chapter discussion of the Team's first evaluation effort, of an unsuccessful attempt, and of a strong evaluation. Appendix B compiles the complete set of all 14 evaluations that derive a cumulative "scorecard" in Table 7. For each outcome, I determine judgments of performance and quality of Team evaluations by using the metric of symbols shown below.

Table 1   Key to logic model "scores"

Symbol	Meaning
✓	Element has been analyzed adequately.
✓+	The Team did an exemplary job in this task
( )	Element is unresolved or conditional.
-	Element or task was not addressed sufficiently.

In meta-evaluation, the logic model gets filled in with symbols denoting a qualitative judgment. The resulting model may look like the example figure below.

Figure 4   Example meta-evaluation of an evaluation process on one outcome
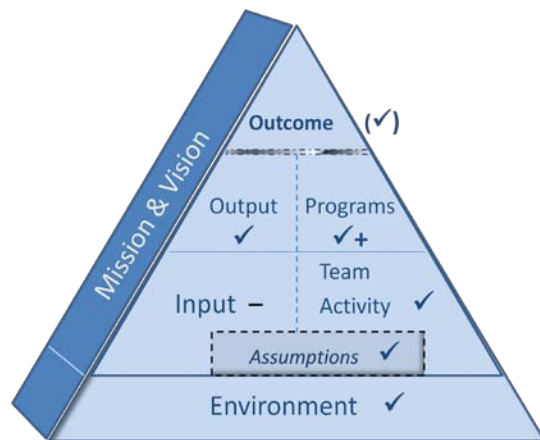
In this example, the researcher has found that the Team used sound assumptions (check), conducted or cited a sufficient external scan (check), located an exemplary wealth (check plus!) of programs related to this outcome, offered a report as output (check), and found preliminary basis for the outcome to be achieved "(✓)". (I use symbols in combination, where appropriate.) However, the parentheses indicate that the outcome has not actually been achieved yet. The

minus sign in the "input" slot informs the project leader of a lapse in process involving student needs and characteristics. Perhaps the Team did not conduct a needs audit to determine what students want to achieve in programs related to this outcome, or measured participation levels were not conclusively linked to demographic factors, mission, operational costs, or other key factors. More detailed narrative would explain what was missing from the Team's evaluative effort with inputs.

My underlying premise is that retro-fitting elements of the case study into a series of logic models and related narratives achieves my research objective of meta-evaluating the campaign presented in this case study. Using this logic model heuristic, analysis weighs what the university system depicted in case study has accomplished. This logic model "roadmap" may also speculatively guide the profiled system's future phases of assessment and possible success in achieving outcomes still pending. In the spirit of meta-analysis, my hope is that readers find descriptions and lessons to compare and contrast to their own evaluative practices, adding to knowledge and consensus about current process and program theory for university student affairs assessment campaigns.

2.0   REVIEW OF THE LITERATURE

This chapter commences with brief arguments for the strategic merits of assessment and of student affairs. Historical origins of student affairs, its justification under student development theory, its emphasis on programming, and its gradual acceptance as one way a campus can ensure its students a well-rounded existence result in a fundamental premise that student affairs is indeed worthy of the rigors of evaluation and assessment (Upcraft & Schuh, 1996; Schuh & Upcraft, 2000).   Accepting this evolution as a foregone conclusion, the literature review addresses topics that student affairs planners and evaluators should consider as they apply or construct a logic model to analyze and possibly guide the assessment cycle. These topics include the political environment of assessment, buy-in among faculty and other stakeholders, standards of assessment, and the evaluator's role in building consensus and mutual trust. I conclude this chapter with origins and application of logic modeling, culminating with creation of the actual heuristic used in this study.

2.1   EVALUATION & ASSESSMENT OF STUDENT AFFAIRS

The literature has investigated the virtues and nuances of student affairs (Rodgers, 1990; Benjamin, 1994; Nuss, 1996; Thelin, 1996; Andres & Carpenter, 1997; Evans, Forney, & Guido-DiBrito, 1998; Komives, 2003), of evaluation and assessment (Cronbach, et al., 1980; Worthen

& Sanders, 1987; Patton, 1990; Eisner, 1990; Scriven, 1991, 1997, 2003; Sanders & Joint Committee on Standards for Educational Evaluation, 1994; The American Association for Higher Education (AAHE) Bulletin, 1995; Gredler, 1996; Chelimsky, 1997a, 1997b, 1997c; T. Cook, 1997; Weiss, 1997; House & Howe, 1999; Rossi, Freeman & Lipsey, 1999; Tananis, 2000; Fitzpatrick, Sanders & Worthen, 2004; W. K. Kellogg Foundation, 2004; Popham, 2008), and of strategic planning (Patton, 1997; Weiss, 1997; Nichols & Nichols, 2000, 2001; Popham, 2008).

Researchers who have explored the intersection of these topics – the history of evaluation and assessment of student affairs as an important objective in university strategic planning – include Worthen & Sanders (1987); Gredler (1996); Weiss (1997); Kellaghan, Stufflebeam, & Wingate (2003); and Driscoll (2006). Upcraft and Schuh (1996) and Schuh and Upcraft (2001) crafted definitive touchstones for theory on assessment in university student affairs. Upcraft and Schuh (1996) contend, "Without assessment, student affairs is left only to logic, intuition, moral imperatives, goodwill, or serendipity in justifying its existence" (p. 11). These co-authors issue warnings on what assessment should not and cannot accomplish: a "quick and dirty" catch-all, such as a single survey; an excuse to point a finger during cost-cuttings; or crisis management. Weiss (1997) admonishes against evaluation as 1) the postponement of important decisions or management ducking responsibility by letting evaluation make a decision for them; 2) positive public relations or a vanity project; 3) window dressing for a decision already made, and 4) fulfillment of a grant request (p. 22). Instead, assessment is a comprehensive and cyclical process that involves all stakeholders[6].

---

[6] Defined by Scriven (1997) as "someone who has made a significant investment in the program, either financial or psychological" (p. 482). This list includes students, student affairs staff, other administrators, faculty, governing boards, legislators, accreditation services, graduates and alumni, funding agencies, and the general public (Upcraft &

Stakeholder needs and possibly conflicting viewpoints are one illustration of how assessment might arise from political origins. Weiss notes, "People who launch it will likely be situated at the top of the hierarchy" (p. 21). Weiss assures that it is important to find out who initiated the idea and why. Even with honorable intents, evaluations might face unfavorable conditions. If a program has few routines and little stability, the process can veer, wander and end up evaluating an obsolete program (p. 24). Other obstacles might include a lack of consensus on programming purposes, lack of essential buy-in, stringent limits on project scope, and prohibitive opportunity costs for resources or staff.

Patton (1990) maintains that evaluation design is exemplary when the study is set up to provide precisely the info needed by program director to make the decision needed. Results are understandable, credible and relevant (p. 149). This study uses a logic model to determine the extent that an assessment team has accomplished those ideals.

2.1.1 **Accountability**

On levels both philosophical and actionable, the purposes of assessment involve proper assignment of accountability (Rossi, Freeman & Lipsey, 1999, p. 204). Weiss (1997) justifies emphasis on accountability by four counts (p. 120):

1)   Requirements for explicit measures of accomplishment summon logic and reason; therefore, evaluation should report on actual results.

Schuh, 2000) . To the authors' list we should add the most belligerent constituency, by Student Affairs colleagues' testimony: parents (Lipka, 2005; Coburn, 2006; Lum, 2006). More recent studies have shown that the presence of "helicopter parents" may help students thrive (Lipka, 2007; Hoover, 2008). An evaluator of student affairs programs cannot ignore competing needs and claims of all these stakeholders.

2)   Agreed-upon standards promote collaboration in an organizational culture. Forthcoming discourse and my case study will demonstrate how a unified organizational culture is one goal of student affairs, and an admirable reflection of mission and vision.

3)   Assigning accountability results in diminished need for centralized micromanagement and rigid rules. Demonstrating improvement and flexibility generate appropriate and responsive types of service. My case study will also show how a student affairs unit effectively delegated tasks.

4)   Accountability can restore [public] confidence and support. This has been a long-term goal of the profiled student affairs unit.

Weiss assures, "Emphasis on goals and accomplishments is in tune with the times" (p. 120). Goals must be specific and clear, translatable into operational terms. However, accountability in student affairs amounts to more than keeping students happy outside the classroom. Conveying this sentiment, Terenzini & Upcraft (1996) claim, "Perhaps the most intimidating question posed to a student affairs practitioner goes something like this: 'Sure, the students like your programs and services, but what evidence do you have that what you are doing is making a difference?' " (p. 217). One danger of a hasty reaction is assessment that emphasizes outcomes alone, perhaps out of context, but perceived as a potential panacea. To strike the delicate and critical balance of allowing students a good time en-route to becoming responsible, productive citizens is arguably the whole purpose of student affairs – and a nod in the direction of student development theory cited as foundation for this study. Skillful assessment can serve desired outcomes at different stages in a student's intellectual and social development. Outcomes fit into a larger "equation" or evaluative process flow that a powerful logic model seeks to depict.

2.1.1.1 Accreditation

Sometimes considered outside the scope of student affairs programming, accreditation deserves mentioning as 1) an environmental / administrative component that shapes and frames assessment cycles on the American campus, and 2) a parallel impetus for formal standards that have evolved in the evaluation field.

Belle Wheelan[7] (2008) offered wisdom pertinent to this topic with the following points about accreditation in higher education. First, accreditation is voluntary: "Nothing says you have to be accredited, if students don't need financial aid." However, accreditation does not serve nor deflect a punitive process, but is designed to look at real data and changes based on mission. Wheelan likened accreditation to the "good housekeeping seal of approval" for the public, proof that a college spends money wisely. Second, accreditation generally follows the three A's of the Spellings Report (approved 2006): access, accountability, and affordability. Wheelan believes that the Spellings Report disassembled negotiative rulemaking, whereby colleges were lured into implementing changes only if new laws were passed. Instead, accreditors are now independent organizations that promote partnership and compromise (Wheelan, 2008).

Tying assessment to accreditation (historically a "burden, an expensive and strenuous routine" (Driscoll, 2006, p. 8)) can save steps and ingrain sound practices of moving beyond data collection and analysis to use of the information. Wheelan said that assessment is linked to accountability – another arguably disagreeable concept, at least where equated with "blame". Such presumption might degrade a positive image of assessment into merely a vehicle for official status and concordant funding and reputation. Still, planners know that accreditation is an

---

[7] President of the Southern Association of Colleges and Schools (SACS) Commission, former Secretary of Education of Virginia, and keynote speaker at a recent education conference in Pittsburgh.

inescapable "high-stakes event" (Fetterman, 1997), and therefore they might most efficiently tie this "burdensome" event to sound, reliable, ongoing practices.

2.1.2 **Evaluation guidelines**

It would be misleading to assume that accreditation incentives have triggered broader development of evaluation standards, even if accreditation incentives might prod new and revised guidelines and standards. Instead, from dialogues have emerged numerous styles and philosophies of evaluation, along with networks that referee and standardize practices. Later in this chapter I will discuss the evaluator's ethical responsibilities, as distinct from but related to standards that may not be enforceable without formal contracts. Standards and guidelines are worth mentioning in terms of how well the case study has adhered to accepted principles of assessment.

Guidelines are a basis of meta-evaluation, which is an evaluation of an evaluation (Patton, 1990, p. 147; Sanders, et al, 1994, p. 185). According to Worthen & Sanders (1987), meta-evaluation was first formally discussed in the 1960's. Development of standards began in 1975 under the direction of Daniel Stufflebeam at the Evaluation Center at Western Michigan University. The thirty standards fall into four "attributes": utility, feasibility, propriety, and accuracy (Sanders, et al, 1994, p. 5). Authors distinguish standards[8] from criteria[9]. "Taken as a whole, the 30 standards provide a proposed working philosophy of evaluation" (p. 17). A profession-wide Joint Committee provides ongoing guidance and authorization. The Evaluation

---

[8] Defined as "principle[s] mutually agreed to by people engaged in a professional practice, that, if met, will enhance the quality and fairness of that professional practice, for example, evaluation" (Sanders & Joint Committee, 1994, p. 2, italics removed).
[9] Defined, in curiously circular or perhaps spiral fashion, as "a standard by which something can be judged" (Sanders & Joint Committee, 1994, p. 205). That "something" can evidently be another standard.

Research Society (ERS) separately published standards on program evaluation (Worthen &
Sanders, 1987), until merging with the Evaluation Network to form the American Evaluation
Association (AEA).

AEA states on its website that the *Guiding Principles for Evaluators* (1994, 2004) serves
as "the cornerstone of good evaluation practice. Developed in 1994 as guidelines for sound,
ethical practice, they have been... reviewed and revised at regular intervals, including most
recently in 2003, in order to ensure that they remain current with the field". I think that the
metric of check marks and other symbols I employ in forthcoming analysis via logic modeling
conforms to standards cited in the table below.

Table 2   AEA Guiding Principles

Systematic Inquiry
Competence
Integrity/Honesty
Respect for People
Responsibilities for General and Public Welfare

AEA (2004) states in the preface of the *Guiding Principles*,

> The five principles proposed in this document are not independent, but overlap in many
> ways. Conversely, sometimes these principles will conflict, so that evaluators will have to
> choose among them. At such times evaluators must use their own values and knowledge
> of the setting to determine the appropriate response. Whenever a course of action is
> unclear, evaluators should solicit the advice of fellow evaluators about how to resolve the
> problem before deciding how to proceed (p. 1).

I think that this interesting disclaimer implies that overriding guiding principles are
mutual trust, instinct, and collaboration – incontrovertibly admirable attributes, as are the guiding
principles themselves. While memory serves that the Team functioned cordially as evaluators,
closer analysis via logic modeling will measure the Team's performance more concretely.

Weiss (1997) cites the importance of "accepted performance standards" in their ties to
quality (p. 132). House and Howe (1999) explain that some guidelines have been derived from

the evaluation community, shared concepts and practices, models and data collection procedures. If elements fit together, the authors recommend no need to go further if not for conflicts (pp. 28-29). House and Howe continue, "Considerations on quality and utility are based on substantive concerns, on specific content rather than formal procedures. The evaluator must make judgments, like a referee in a ball game must follow sets of rules, procedures, and considerations" (p. 29). Echoing the work of Robert Stake, they say judgment comes first, standards later (p. 30), tying into accountability[10].

I have observed that much evaluative activity takes place under less formal, more intuitive guidance[11]. Upcraft and Schuh's (1996) opening chapter lists principles to guide effective assessment. These authors' list seems a valid "mission statement" or checklist for assessment, all of whose elements, the authors claim, must be present for the time and investment to be worthwhile.

- Student affairs assessment begins with educational values.
- Effective assessment reflects an understanding of organizational outcomes as multidimensional, integrated, and revealed in performance over time.
- Assessment requires clear, explicitly stated goals.
- Assessment requires attention to outcomes, but just as importantly to processes.
- Good assessment is ongoing, not episodic.
- Representatives from across student affairs and the institution should be involved.
- Assessment makes a difference when it begins with issues of use and illuminates questions that people really care about.
- Assessment is part of a larger set of conditions that promote change.
- Through assessment, student affairs practitioners uphold responsibilities and expectations of students, the institution, and the public (pp. 22-24, borrowing from AAHE, 1992).

---

[10] Patton (1990) opposes (or did at the time) universal, standardized evaluation measures in belief that local program processes are too diverse and client outcomes too complex to be represented nationwide, or even standardized by some narrow set of prescribed measures and methods, akin to a state religion (pp. 493-4).

[11] Accordingly, the "broad legislative framework for many … current reform efforts" under the Government Performance and Results Act (GPRA) of 1993 (Wholey, 1997, p. 125) probably represents overkill by standards in a student affairs context.

As assessment is influenced and sometimes dominated by federal, local, and institutional politics, accepted standards are part of a complete logic model's "environmental" component. My case study will examine moments of interplay between judgment and standards. The next sub-chapter discusses how the evaluator might navigate such murky waters.

2.1.3 Roles of evaluators

We live in a political world.

Bob Dylan (1989)

Weiss (1997) characterizes evaluation as "systematic assessment of operations and / or outcomes of programs or policy" (p. 320). The next paragraphs describe how an evaluation specialist or team accomplishes tasks from a multitude of approaches in a politicized context. This sub-chapter a) explores vantage points from which evaluators might draft workable logic models; and b) gives foundation for how effectively the Assessment Team in the case study performed its evaluative role.

Weiss continues,

Intrinsic to evaluation is a set of standards that (implicitly or explicitly) define what a good program or policy accomplishes and looks like. The purpose is improvement of programs or policy by encouraging elimination of unsuccessful interventions or giving guidance for how existing intervention can be modified (p. 320).

House and Howe (1999) explain:

In general, evaluators are responsible for investigating the pertinent body of social research against which programs may be understood and compared. Evaluators are also responsible for interpreting such research and judging its merits. In doing this, they are required to be objective or unbiased [italics removed], not in the sense of refusing to offer judgments of their own, but in the sense of grounding such judgments in defensible

methodological and moral principles. Rarely will programs or policies be so innovative that no pertinent research exists… (p. 125).

A wealth of "pertinent research" documentation justifies the utility of the common "external scan", part of any strategist's toolkit. House and Howe (1999) say that evaluators' "strenuous efforts" and conformance with democratic principles provide "a much better basis for public understanding and decision than would be available in the absence of evaluation…. Average citizens can no more collect data and conduct their own evaluations than they can construct their own flu vaccines" (pp. 129, 132)[12].

According to Patton (1990), what stakeholders expect of evaluators is 1) "to confirm what they know that is supported by data; 2) to disabuse them of misconceptions, and 3) to illuminate important things that they didn't know but should know. Accomplish these three things and decision makers can take it from there" (p. 423). As I can attest from experience, the evaluator can act as a trusted consultant – which may be, according to Scriven, a ceremonial and misleading designation. Scriven (1997) is careful to separate an evaluator's role from consultant as service provider. He calls the evaluator instead "an expert witness…, bound by exactly the standard oath" of the whole truth (p. 496).

In the process, evaluators "search for totality – the unifying nature of particular settings. This holistic approach assumes that the whole is understood as a complex system that is greater than the sum of its parts" (Patton, 1990, p. 49). I see that this notion of totality applies to the context of student affairs, as more than a network of programs and advisors and administrators. Unity under mission and vision seeks to ensure that students experience housing, dining, and programming seamlessly, with transparent resolution of problems, regardless of unit-level

---

[12] House and Howe do not proceed to a depiction, perhaps directed by a Frank Capra acolyte, of a dim parallel world, void of evaluators and restored upon their return.

accountability and funding. These intangible outcomes seem difficult to measure, although a detailed logic model might point towards qualitative indicators.

Regarding an evaluator's level of participation, House and Howe (1999) say that guides of conduct might be an experienced evaluator's own intuition or an explicit conception of public interest. In not championing a particular view or interest, their stance stops shy of advocacy. House and Howe argue that the evaluator should advocate only for democracy and public interest, not directed outcomes, nor should evaluators play role of neutral facilitators among advocates[13] of competing "value summaries" or stakeholder "constructions" (pp. 95-6). Instead, as House and Howe (1999) state right up front in their introduction, "The evaluator role is compatible with democracy" (p. xv).

Patton (1990) warns against adverse effects of an evaluator's actions: 1) bias introduced as subjects' behavior changes under an evaluator's watch; 2) change in the evaluator's behavior or instrumentation approach; 3) predispositions or biases of the evaluator; and 4) evaluator incompetence, in terms of training or preparation (p. 473). Patton confesses his own view is that "evaluator effects are considerably overrated, particularly by evaluators" (p. 474). As remedy, Patton recommends, "Therefore, take it seriously, but in a balanced approach" (p. 475).

So, the evaluator serves an important role in cool balancing between detached methodological consultancy and rabid cheerleading for programmatic innovations. The evaluator's duty to carry the political tone does not mean (s)he stands aloof and isolated, peering alone from a hilltop over a smoldering horizon. As the evaluator is accountable for the condition of that horizon, the next two sub-chapters explore dialogues on the politics of evaluation and on

---

[13] For more on advocacy evaluation, see Jennifer C. Greene (1997) "Evaluation as advocacy", *Evaluation Practice*, 18: 25-35 and Melvin Mark, et al, (2000) *Evaluation*, viewing evaluation through the lens of social betterment. Pundits could argue whether this grand concept is out of scope for student affairs, which seeks only to make a confined domain better.

buy-in, topics that deserve inclusion in the "environment" and "impact" portions of my logic model.

2.1.3.1 Politics

IHE units have created new streams and logics of organization reporting, documentation, and analysis to justify costs, boost efficiency, and comply with standards[14]. As the traditional campus is not insulated from accountability, planners have also found value in enlisting evaluators to gather evidence that supports directives and avoids conflict between competing stakeholders. I concur with Tananis' (2000) depiction of the evaluator as an "insulated expert" entrusted with keeping decision makers informed of shifting trends and with Worthen & Sanders' (1987) definition of the "expertise" evaluative schema. However, Schwandt (1997) notes distinctions between "typical" consultancy and "evaluation as practical hermeneutics" on three grounds (p. 79). First, dialogue and reflection deflect the "pedagogical authority inherent" between consultant / experts and clients. Second, a humanizing tradition of education and culture informs dialogue and engagement that lead to problem solving. These sentiments echo discussions by House and Howe (1999) on the dialogical approach and by Patton (1990) and Weiss (1997) on power dynamics, but contrast with Eisner's (1990) idea of connoisseurship. Third, Schwandt claims, citing William May (1992) on "critical intelligence":

> The general logic of evaluation is suited to generating operational intelligence in clients. This kind of intelligence is instruction on the status of means and means-end reasoning; it is directed at helping a client get to there from here. Critical intelligence, on the other hand, is the ability to question whether the there is worth getting to" (p. 79).

---

[14] These trends may be especially true among profit-driven institutions, such as on-line and virtual learning modules. Evidence exists in a large number of job postings for evaluators and directors of institutional research, demonstrating how two-year and community colleges appear eager to enlist evaluation for accreditation purposes and credibility. These topics could be ripe for analysis, with a hypothesis that other types of colleges differ on accountability grounds from the traditional four-year campus depicted here, where an active student affairs presence plays a greater role in the development of the whole student.

I think that these two types of intelligence add up to institutional self-empowerment, which is what Pitt Student Affairs has aspired to achieve particularly in latest phases of its assessment campaign.

In sum, I see that evaluation and assessment often involve politics. Elements of high cost, time consumption, choices, impact and visibility are attributes that define a political process or event. Evaluation and assessment can provide evidence for policy change, while serving as an ostensibly benign method for academic leaders to (re-)allocate resources.

However, my impression of political components of a malleable logic model is that politics cannot be the lone or primary driver of assessment. While it may be argued that all assessment takes place within political constraints, justifying assessment as strictly a political function or requirement may compromise its aims and outcomes. Especially in a humanistic, service-oriented domain like student affairs, political posturing has a tendency to trickle down. Let data speak, rather than agendas. Rather, politics must be recognized as part of an environmental scope as practitioners conduct assessment and make appropriate recommendations, based on evidentiary data and analysis, to drive subsequent phases of politicking in student affairs or other university strategic units. My case study portrays a department that aspires to conditions where political nature and origin fade when assessment has become part of a natural, expected administrative flow. A good logic model can help identify and contextualize those political factors. Thus my approach is to include politics in the "environmental" framings that filter knowledge into "assumptions" and "activities" slots, without calling singular attention to politics. This notion does not overstate logic modeling as more than an instrument used by people to clarify political context also created by people (Tananis, 2009, unpublished).

2.1.3.2 Buy-in and consensus

I have seen that a challenging role for evaluators is to achieve buy-in – arguably a subset of politics – among staff and faculty skeptics for programs and for the evaluation processes that attempt to improve those programs in light of competing costs and directives. This sub-chapter addresses how an evaluator balances conflicting aims and findings to achieve acceptance from stakeholder communities.

As Patton, Weiss, House and Howe, Eisner, and Upcraft & Schuh devote large chunks of narrative to this topic, the Joint Committee (1994), too, grants an entire chapter to responses to evaluation. "Evaluators must not assume that improvements will occur automatically once the evaluation report is completed" (Sanders, et al, 1994, p. 59). An important consideration is to gauge the receptiveness for rigorous study. An assessment specialist or team may be disheartened when a well-conducted study is not put into action. One cause could be faulty communication between evaluators and the stakeholder client. I have seen this happen and have witnessed a lowering morale. On the brighter side, I have also seen the glow of satisfaction among team members and clients upon successful program implementation or customization based on evaluative findings and recommendations.

First, the warnings: Weiss (1997) says that if evaluation is "foisted upon a crew, confrontation and suspicion can compromise salability" (p. 101), especially for programs facing summative judgment; can infringe on autonomy; and can result in power imbalances. The evaluator can take action to boost the probability of buy-in, if not transition towards a community that welcomes regular, ongoing assessment and evaluation.

Chelimsky (1997b) ranks credibility first among pivotal political implications of evaluation, along with courage, dissemination of strong findings, linkage across disciplines and with basic research, access to information, and "more realistic training of evaluation students" (pp. 58-64). House and Howe (1999) say that the two-sided nature of politics dictates that the conflicting side gets "buried or ignored in favor of sentiment" (p. 93). The authors find utility and balance in including conflicting values, and they claim that some researchers have dodged a thorny issue by labeling activities "critical" rather than taking an "apolitical" post-modern stance on education (House & Howe, 1999, p. 78).

Two reasons that "political constraints inhibit attention to evaluation" are that 1) evaluation deals with real-world programs created and maintained through political processes; and 2) people in higher echelons of academic government make decisions about programs (Weiss, 1997, p. 312). House & Howe (1999) state the corollary that people without power rarely sponsor evaluations. In student affairs, presumed beneficiaries of programs rarely do, either. This dilemma invokes the ethical question of who owns findings, such that a sponsor might exercise power to reshape a report in ways that meet his own needs (House & Howe, 1999, p. 119). Chelimsky (1997c) concurs that "excessive preoccupation with the acceptability of findings risks turning evaluations into banal reiterations of the status quo" (p. 106).

Since programs do not simply happen but come into being through political processes, program directors can view evaluation as more threat than help. Weiss (1997) says that a political statement is implicit in selection of some programs to undergo evaluation and others to escape scrutiny (p. 313). Program evaluation thus faces a double layer of skepticism, one against evaluation itself and another for or against program "ownership" or sponsorship among opportunity costs. This case study involved moments of head-butting, and I will show how a

logic model can help categorize competing interests such that ranges of opinions are represented in deliberation and in decision-making.

According to House & Howe (1999), evaluations secure acceptance by meeting three explicit requirements: inclusion, dialogue (authenticity), and deliberation (validity and expertise) (pp. xx, 11-12). Evaluators should not ignore imbalances of power nor pretend that stilted dialogue is open and inclusive. House and Howe observe in power-driven evaluations a contrast to "emotivist" or preferential democracy, a philosophical bent that fits nicely within student affairs sensibilities. Weiss (1997) agrees with emphasis on knowing "the reality of the program over the rhetoric. Research questions involve net effects of program over and above what would ordinarily happen in its absence" (p. 322). This notion opts for causality over politics as the driver towards intended outcomes.

Patton (1990) might criticize the idealism of House and Howe's conception of dialogic democracy in evaluation:

> Experience suggests it impossible or at least impractical to have the same degree of closeness or distance from every group or faction. Evaluators, human beings with their own personalities and interests, will be naturally attracted to some people more than others. To resist those attractions may hinder observation and integration.

> Perhaps the most basic division is between staff and participants, between those who are paid for their responsibilities and those who are recipients of program offerings (p. 262).

To level this playing field, Patton recommends "cultivation and use of key informants, done carefully as not to arouse political hostility or personal antagonism. Key informants can provide particularly useful information about what is happening in subgroups to which the observer does not or cannot have direct access" (pp. 263-4). The Assessment Team profiled in this case study used RD's as trustworthy agents in this capacity.

Even with a masterfully crafted report, Weiss (1997) notes that policymakers will not likely use evaluation results as a single basis for their decisions. A study might reveal unattractive results whose visibility may not prima facie lead to a program's termination or cutback. Weiss and Patton assure that fears of immediate negative consequences from evaluation are usually unwarranted. However, obligation to protect interests of people in the program can conflict with a responsibility for honest reporting (Weiss, 1997, p. 109).

Husen (1997) concurs with policy perspectives that indicate a comprehensive and useful assessment. Overall social and economic conditions determine educational policy, but campus policy must link to individual, customized needs for students. Accommodations for handicapped people are one example. Campus operations build ramps not because the government ordains this accommodation (the cynical side of the same coin, tails), but because even one student requires access to her classroom and dormitory (heads!). This element corresponds to the "input" block of my logic model.

While consensus may not be immediate, buy-in may proceed in what Driscoll & de Noriega (2006) call a "ripple effect" (p. xv). While these authors' observations took place in the context of accreditation, the same effect can happen in student affairs programming with open communication and enthusiastic marketing of ideas. Good ideas spread logically and organically.

2.1.3.3 Data gathering: interviews and focus groups

As buy-in, dispute resolution, and consensus depend upon persuasive evidence, evaluators are entrusted with collecting reliable and valid data (Patton, 1990; Weiss, 1997; House & Howe, 1999; Upcraft & Schuh, 1996). The Team relied upon internal scans, external scans, and interviews / focus groups for most of its data gathering. This sub-chapter summarizes

sound principles of gathering data through direct interaction with students and other Student Affairs stakeholders.

Patton (1990) explains,

> The purpose of interviewing is to find out what is in and on someone else's mind, not to put things in someone's mind but to access another perspective, to find out things we cannot directly observe. Qualitative interviewing begins with the assumption that the perspective (sic) of others is meaningful, knowable, and able to be made explicit…. The task for the interviewer is to make it possible for the person being interviewed to bring the interviewer into his or her world (p. 278-9).

The Team presumed that interviewees accurately represented campus consensus. On cue, Patton (1990) tells evaluators to "beware [such] presuppositions, a usual, natural part of interactions, useful in assuming respondent has something to say (but leave it at that, i.e., 'Did you go' before asking 'How did you like it?')" (p. 303).

Focus group interviewing was developed (in 1950's) in recognition that many decisions are made in a social context. Checks and balances among participants can weed out extreme views (Patton, 1990, p. 335-6), much like tossing away the minimum or maximum values of a quantitative data set that would otherwise skew a curve. Patton (1990) notes common characteristics of and conclusions about interviews and focus groups:

- Staff interviewers reduce timing and costs (p. 344). Outsourced consultants would probably do a fine job, and might exercise well-honed authority and credibility, but at costs that a modest Student Affairs unit might not afford. Instead, Student Affairs practitioners had 1) a closeness and credibility that allowed them to relate to participants, and 2) familiarity with settings that enabled them to blend in with minimal disruption to ordinary work flow and reporting channels.

- Patton voices an intuitive reminder against conjuring. "Guessing the meaning of a response is unacceptable; if there is no way of following up, then areas of vagueness become

29

missing data" (p. 353). This principle is consistent with the purpose of research interviews "first and foremost to gather data", not to change or judge people, and not to offer therapy (p. 354). Patton identifies rapport (personal) and neutrality (content) gained in not reacting to feedback (p. 316) – a journalistic approach. In Student Affairs context, evaluators avoided joining in a homesick freshman's tears, deferring to RA's to satisfy that function. With too much empathy, an evaluator might conclude from such encounters, "Yeah, this place is sad."

- Interviews and focus groups are a rich source of quotations to be used in reports to "illustrate the power of qualitative data, beyond a concept or label to tie together data" (p. 392).

- Other guidelines offered by Patton (1990) are that leading questions can bias answers; a two-way flow can result in reliable data; reinforcement cues inspire natural reactions among respondents; and thanks are appreciated, if not expected. For these protocols, I confess that I presumed their observance, rather than strict monitoring or supervising of fellow Team members who conducted interviews. A valuable lesson I learned about team morale was to defer to Team members' professionalism and decorum – trusting colleagues to perform tasks to the best of their abilities.

Both "informal interviews and more formal indicators reflect the reality of program achievement, just different aspects of it" (House & Howe, 1999, p. 21). I believe that interviews and focus groups have particular utility in Student Affairs, given the unit's continual proximity to students whose voices should be heard.

2.1.3.4 Benchmarking

In my opinion, a primary task, moving forward, of the Assessment Team portrayed in this study is to transform data gathering and analysis into benchmarking, a powerful way to achieve continued buy-in. For data to be purposeful, and for recommendations to deserve consideration as potential policy, stakeholders often require measurable backup. Benchmarking should be included in the "activity" layer of the logic model used here.

Schuh and Upcraft (2000) define benchmarking as a comparison of current status to "best practices" on level with internal (i.e., campus wide) standards, among competitors or peers, or functional / generic, "comparing performance across organizations" (p. 332). Benchmarking allows an institution to set target dates and milestones for a complete assessment process, including affordability / feasibility study. Benchmarking encompasses multiple layers not limited to a step-by-step attainment of the next target, nor a before and after snapshot, but also a forecasting of durable programming to achieve certain goals.

My pilot test case studies and activities of the Student Affairs Assessment Team have revealed that University of Pittsburgh has done a commendable job in its strategic plan to view universities such as Delaware, Syracuse, Penn State, and Buffalo as peer institutions, and Michigan and UNC-Chapel Hill as "aspirant" institutions. Aspirant or peer status derives from a combination of profile factors, including the collective talent of a student body and incoming freshmen, endowment, athletic philosophy, graduate programs, systems of housing, post-graduation trends, and alumni activity. Effective benchmarking can thus indicate how Pitt can attract a talented student body that competes with other regional schools, while possibly drawing scholars who might otherwise attend fine state universities in Michigan, Wisconsin, California,

Pennsylvania and other well-acclaimed systems. These lists can be included both in "environment" and "assumptions" boxes of a logic model.

Knowledge of field methods can guide an evaluator to achieve triangulation that accomplishes collaboration and corroboration. Such approaches might include direct contact with project managers conducting parallel studies at peer institutions, phone interviews and surveys, or a focus group among organizational rosters, such as the Council of Alumni Association Executives (CAAE) and the American College Personnel Association (ACPA). A wide spectrum of campuses would allow the planner or evaluator to contrast various tones, communication channels, and ranges of buy-in. I diagnose no shame in first contacting immediate or established leads, under the same logic that an evaluator might employ a "convenience" sample for a survey.

Upcraft and Schuh (1996) offer disclaimer that "benchmarking should be done correctly or not at all" (p. 250). They may imply that benchmarking is not essential for project success, which could rest on its own merits. That claim would strike me as incidental, while triangulation is advisable for concurrence and efficiency.

I think the most important facets are realistic guidelines and comparisons. Benchmarking with quantitative vigor and qualitative sense can strengthen hunches and beliefs into convictions about effective programming and improvements. Benchmarking rescues outcomes from dormancy.

2.1.3.5 Meta-analysis

As credibility builds, the evaluator's efforts can contribute to ongoing dialogue, possible consensus, and theory – both as researcher and practitioner (Frechtling, 2007, p. 6). Weiss (1997) says, "Many qualitative researchers insist that all truths are local and contingent. Each

program is its own world and has to be understood in its own terms…, but individual studies can be pooled with hundreds of other studies of programs of like kind through meta-analysis" (p. 69). Sharing, reciprocity, and collaboration in academic discourse result in meta-analysis, which Weiss (1997) defines as aggregate results of evaluations of similar programs (p. 133). Patton (1990) notes that "where analysis compares several programs or communities, the inductive approach looks for unique characteristics that make each setting a case unto itself" (p. 45), but that each unique case may contribute to a collective wisdom. Popham (2008) calls meta-analysis "making sense of the myriad" and "readily understood quantitative tactics for coalescing the results of dissimilar investigations" (p. 18).

Weiss (1997) elaborates, "A major advantage of meta-analysis is to improve estimation of outcomes of a particular type of program" (p. 244). Pooled information results in cost-benefit, -effectiveness, -minimization, and -utility, since deflected or "shared" costs are lower than the expense of one large study (p. 244). Pooling can also increase an overall sample size, and improve the precision of estimates of program effects, of ranges, and of historical periods. These benefits improve both internal validity (soundness of relationships between inputs and outcomes) and external validity (generalizability) (p. 241-2). Popham (2008) concurs, "Skilled meta-analysts can synthesize the results from a variety of studies so that, despite study-to-study differences, we can still derive generalizable conclusions from a welter of seemingly different investigations" (p. 18). Thomas Cook (1997) puts this observation into a social utility context: "Perhaps the most striking discovery with meta-analyses is how frequently we find that interventions have positive impacts on lives" (p. 37). As Cook notes that those impacts are "often less than we would like" (p. 37), Weiss notes a distinction between "evaluation synthesis" that

33

happens when first-level evaluation has direct exposure to program workings, such that effects of meta-analysis might be remote (p. 243).

Weiss claims that meta-analysis contributes state-of-the-art knowledge to forecasts (p. 283). Perhaps what makes pooled knowledge timely and interesting is an element of creativity. Patton (1990) advises the evaluator to aim for "both the science and the art" of this craft, to resist the usual grounding of evaluation in critical thinking at possible expense of "creative imperative of our work" (p. 434). As Chelimsky (1997b) proclaims, "Our job… is not revolving but cumulative" (p. 67). One of my objectives in this study is a meta-analytical contribution to ongoing dialogue about evaluation and assessment.

2.1.3.6 Summary of the evaluator's role

I concur with Patton's (1990) declaration that the evaluator is entrusted with "removal or at least awareness of prejudices or assumptions about investigated phenomen[a]" (p. 407). Patton's observations seem consistent with Weiss' (1997) conclusion: "Evaluators do research that has a chance of improving the well-being of people in need. Done well, evaluation is noble work" (p. 326).

This noble work might involve many styles and approaches, each conducive for different contexts and settings. Sometimes an evaluator's role is to move forward with new ideas, whether these ideas might fine-tune a successful process, suggest new schemes, guide towards improvement, or even overhaul programs. During our campaign, our Assessment Team did not know about logic models and their possible utility in analyzing or structuring evaluation activities, targets, and time frames. This study explores a parallel universe where if we did know about logic models, then the actual campaign might have followed such a roadmap. Depending on findings from upcoming analysis, as the Team's former chair and an evaluator by trade, I

could accept a daring role in suggesting logic modeling as a viable approach in the ongoing assessment campaign.

2.2   HEURISTICS

This sub-chapter discusses various "roadmaps" that evaluators and decision makers can use to explain or enhance projects ranging from a single program evaluation to a whole assessment campaign. According to Patton (1990), such models are not so much recipes as frameworks (p. 115). Here I describe candidate research frameworks and conclude with my choice for analyzing this case study.

2.2.1   I-E-O and other acronyms

Upcraft and Schuh (1996) and other researchers cite Alexander (Sandy) Astin's[15] input-environment-outcome (I-E-O) model (1991), a process flow of variables to measurable outcomes (Weidman, 2006, p. 256), as an influential and popular framework. The I-E-O model encapsulates evaluative ideals described in preceding sub-chapters. Identifying and classifying all "I", "E", and "O" factors – along with process flows and causality denoted by the arrow – could signify broad assessment.

---

[15] Yes, he is in fact the brother of John Astin, the actor who portrayed Gomez Addams. Based at UCLA, Dr. Astin served on the dissertation committee of a trusted colleague, who assured me that dinners with both brothers present were unforgettable encounters.

Figure 5 **Astin's I-E-O model (1991)**

$$I \qquad E \rightarrow \qquad O$$

College student inputs include their high school grades, expectations, standardized test scores, preliminary majors, and a "variety of demographic measures" (Upcraft & Schuh, 1996; Weidman, 2006). One assumption is that to measure the institutional effect on students, an evaluator must isolate and control initial differences in inputs. Weiss offers a definition that inputs relate to participants and characteristics; the composition of a group is conceptualized as an input (pp. 134-5). She lists among inputs: organizational auspices; budgetary amount; nature of staff; management experience; location; plan of activities; methods of service; philosophy and principles of operation; purposes; period or span of operations; size; client eligibility; community characteristics (p. 132). I would debate that some of these factors are better described as "environment" and "activity". However, the critical feature is their inclusion within recognition of context.

The environmental component consists of numerous forces that influence how students learn. Weidman (2006) explains, "Colleges and universities are not, after all, encapsulated environments" (p. 258). According to Upcraft and Schuh (1996), Astin identified seven environmental classifications: institution (i.e., enrollment, profile as liberal arts or trade school, endowment); peer group; faculty; curriculum; financial aid; major field choice(s); residential structure; and involvement both in academics and extracurriculars. Astin assumed and recommended a longitudinal approach to collecting and tracking this data, mainly to deal with the complexity of estimating change over time, without limiting analysis to two "snapshots" or a simple before-and-after permutation.

36

Patton (1990) says about environment, the importance of setting is descriptive but not interpretive (p. 219). History plays a part in understanding the social environment (p. 223). I hope that my descriptions of the Pitt campus (forthcoming in chapter 4.1) adequately set the scene and its origins. The nature of a program emerges from the interaction of the physical setting, the social environment of people in the program, the nature of staff leadership and administration, and the activities conducted or sponsored in the program.

Among outcome components of Astin's and many models, the "easiest and most measurable collegiate outcomes are grades and retention" (Terenzini and Upcraft, 1996, p. 222). Other outcomes include intellectual growth, identity, interpersonal skills, values, career development, and quality of life. "Historically, student affairs typically has focused on out-of-class 'non-academic' outcomes…" (p. 222). My case study features fourteen such "student baseline outcomes", measured and analyzed in various stages over two academic seasons.

Thus, Astin's framework for assessment design collects "three different kinds of information on students: what they were like when they came to college, the nature of their experiences while in college, and what they are like when they leave college" (Terenzini and Upcraft, 1996, p. 222). Terenzini and Upcraft (1996) proceed to warn practitioners strongly against "incomplete" outcome models that leave out any of these components or even separate them from a continual flow. "The bottom line is that the I-E-O model cannot be used piecemeal, in spite of the fact that this is frequently done" (Terenzini & Upcraft, 1996, p. 224). I would argue that, from a programming standpoint, the critical element of Astin's model is the humble arrow, which signifies logical flow from inputs and environmental factors to measurable, desired outcomes. Weiss (1997) describes this link as internal validity (p. 241), an attribute my logic model seeks to preserve.

Upcraft and Schuh (1996, 2000) conclude how its simplicity, broad nature, and depiction of the student component of assessment from a psychological and developmental viewpoint render the Astin model a popular option for decision makers to assess student affairs programs. Similarly, CIPP (Context / Input / Process / Product) has been "refined and elaborated by constant use" (House, 2003, p. 10) such that yet another case study embracing this model might not add useful knowledge to ongoing dialogue[16]. Patton (1990) cites a possible weakness of these models: "Simple statements of linear relationships may be more distorting than illuminating" (p. 423). Andres and Carpenter (1997) agree that "parsimonious, unidimensional 'integration' models are limited in scope, and hence do not adequately reflect the complex and multileveled lives of today's students" (p. 28). While I-E-O and CIPP are enticing models for a rudimentary campaign, my thinking has progressed to where I consider logic modeling a more thorough and enduring roadmap for a structured project.

2.2.2 Logic modeling

The terms "logic modeling" and "program theory" refer to a chain of assumptions that explain how program activities lead step by step to desired outcomes (Weiss, 1997, p. 265; Rossi, Freeman & Lipsey, 1999, p. 214-15; W.K. Kellogg Foundation, 2004, p. III). Patton (1990) uses synonymously the terms "program approach" and "theory of action" (p. 202). Taylor-Powell and Henert (2008) agree with Weiss that a logic model is a framework, specifically for "describing the relationships between investments, activities, and results. It provides a common

---

[16] Nonetheless, Stufflebeam's invention enjoys wide application; a recent article lists 41 projects conducted by the Western Michigan University Evaluation Center enlisting CIPP in areas ranging from community development to testing to meta-evaluation, and four involve higher education systems (Stufflebeam, 2003, pp. 58-61).

approach for integrating planning, implementation[17], evaluation and reporting" (p. 54.10). They also warn what a logic model is not: theory, reality, nor an evaluation model or method. Rather, it enables planners and evaluators to visualize a method. The process may begin with a set of guesses, hypotheses, or preliminary hunches (Weiss, 1997, p. 266). Weiss says that qualitative program theory tends to be discovered in the course of study; a logic model emerges from data. I grant that my logic model evolved after years of exposure to this data and the processes that gathered it.

According to Weiss, the most basic program theory model consists of a) program inputs, such as resources and organizational auspices; b) program activities and their implementation; c) interim outcomes, as a chain of responses; and d) desired end results. These components map into the diagram below.

Figure 6   Distillation of Weiss' (1997) sample program theory model (p. 63)
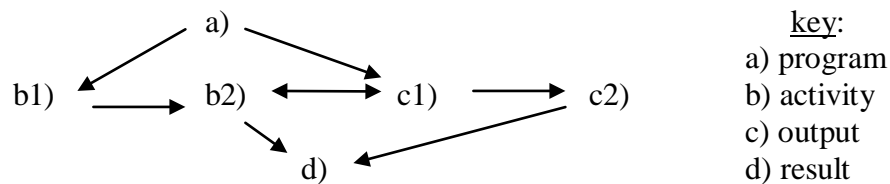
A simple approach might be to sort elements into a matrix with inputs, activities, outcomes, and results as column headers, by Weiss' categorizations. Weiss explains that, by mapping and classifying, "program theory alerts the evaluator to vital issues and opportunities" (p. 71).

The W.K. Kellogg Foundation (WKKF) (2004) concurs on how a visual representation grants "opportunity to chart the course" (p. III) and "share[s] knowledge about what works and

---

[17] Implementation assesses the extent to which activities were executed as planned, since a program's ability to deliver its desired results depends on whether activities result in the quality and quantity of outputs specified (WKKF, 2004, p. 37).

why" (p. 1, italics removed). WKKF offers another simple, linear representation that might remind researchers of Astin's (1991) I-E-O model.

Figure 7   WKKF (2004) diagram on how to read a logic model (p. 3)

Resources / Inputs	→	Activities	→	Outputs	→	Outcomes	→	Impact
①		②		③		④		⑤

Planned work   Intended results
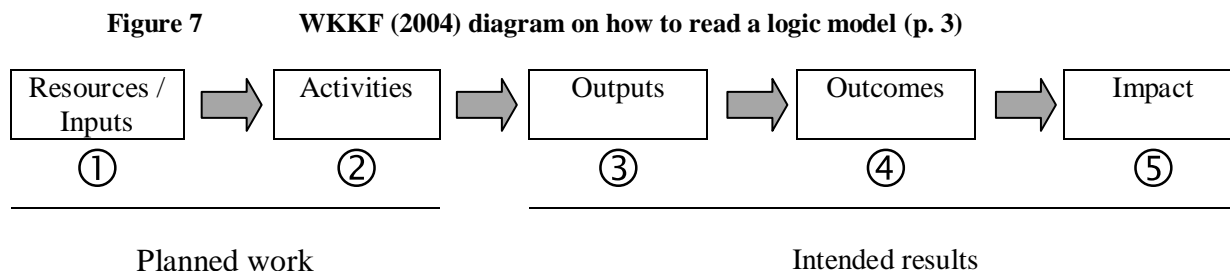
Reading the model left to right, planners discern program basics sequentially, from planning through results. WKKF (2004) supplies definitions of categories that I have used liberally throughout this study and for which I perhaps presumed common understanding. These definitions (p. 2) should deliver welcome consensus:

1. Resources or **Inputs** refer to the human, financial, organizational, and community resources a program has available to direct toward doing the work. (For my diagramming and logic, I concentrate on *students* as sources of input. I conjecture that this approach is consensus in Student Affairs.)
2. Program **Activities** are what the program does with the resources, including processes, tools, events, technology, and actions that are an intentional part of the program implementation. These interventions are used to bring about the intended program changes or results.
3. **Outputs** are the direct products of program activities and may include types, levels and targets of services to be delivered by the program. Frechtling (2007) explains, "Outputs are the simplest and most immediate indicators of the progress of [a program] theory" (p. 24).
4. **Outcomes** are the specific changes in program participants' behavior, knowledge, skills, status and level of functioning. Short-term outcomes should be attainable within 1 to 3 years, while longer-term outcomes should be achievable within a 4 to 6 year timeframe.
5. **Impact** is the fundamental intended or unintended change occurring in organizations, communities or systems as a result of program activities within 7 to 10 years. In the current model of WKKF grant-making and evaluation, impact often occurs after the conclusion of project funding.

Patton (1990), citing his own theories on utilization-focused evaluation (1986), explains how a logic model leaps off a page or whiteboard into practical operation:

> Formulating a program theory of action is a step beyond either implementation or outcome evaluation alone. The program's theory of action links the two. It depicts in concrete terms how inputs and activities are related to outcomes and impacts to constitute a holistic program model or treatment. [There is] contrast sometimes between espoused theories, what people say they do, and theory in use – real priorities and happenings (p. 107).

With cyclical planning, the evaluator adapts and adjusts steps so that pieces fit together into a "cohesive and convincing evaluation" (p. 73). Weiss (1997) also recommends selectivity in data collection to achieve cooperation from staff and students; meticulous analysis and coding, reality checks, tests, modeling, and re-analysis; and sifting from a surfeit of data. Weiss emphasizes the official statement of program goals, which may involve cutting through some rhetoric (p. 75). Staff hunches and observations can supplement official statements. The evaluator must not discount institutional knowledge, especially in a healthy organization. These sources can enable the evaluator to discern between attribution and causality, as favorable results could be derived "anyway and regardless". Weiss explains causality in that program improvement should be traceable to reasons for shortfalls and achievements; therefore, it is necessary to track program theory explicitly and implicitly through evidence (p. 76). In the process, the evaluator must "guard against [any] tendency to mind only intended or supposed effects, as unintended features may also surface", then decide which questions to pursue, based on a) decision timetable; b) relative clout of interested parties; c) stakeholder preference; d) uncertainties / gaps in knowledge base; e) practicalities; f) assumptions of program theory; g) potential use(s) of findings; and h) the evaluator's professional judgment (p. 77).

To trace patterns and causality, Weiss concurs with Richard Elmore's (1996) suggestion to try "backward mapping", where outcomes link directly to origins. Backward mapping

encourages the planner / modeler to restate *intended* outcomes as criteria[18] at a program's onset – hopefully to develop into *achieved* outcomes.

Weiss implies that outcomes may manifest themselves in different phases, that short-term versus long-term measures may show a "sleeper" effect (p. 124). The best way, says Weiss, to reconcile a delay or transition between complementary outcomes and to conceptualize interim indicators, or benchmarks, is to consider the program's theory of change (p. 124). Weiss lists among possible indicators the type of programmatic activity; characteristics of staff offering services; frequency and strength of service; duration; intensity; integrity of design; size of group; stability / shift in focus; quality of service; and responsiveness to individual needs (p. 130). An important feature is that all of these indicators are measurable in some way – mathematically, on a gradation, by comparison to peers, by point-A to point-B snapshot, or by qualitative judgment.

A logic model allows the evaluator to place and contextualize that causality, what WKKF (2004, p. 7) and Rossi, Freeman & Lipsey (1999, p. 238) imply in "if… then" chains of reasoning. Patton (1990) says that causality can be tested in randomized experiments, but that there is an "important distinction between relative degrees of calculated manipulation." Causality typifies a "naturally unfolding program" (pp. 42-3, italics removed). Rossi, Freeman & Lipsey (1999) also distinguish between "perfect" and "good enough" reckonings of causality (p. 239-40). The figure below illustrates causality in program theory (Weiss, 1997, p. 129).

---

[18] Eisner (1990) says that criteria should be based in triangulation or what he calls structural corroboration, the next best quality to being watertight (p. 55). Eisner credits Dewey in *Art as Experience* (1934) for distinguishing criteria, as function of judgment, from standards based upon physical attributes (p. 101).

Figure 8 **Program causality**

	Program		Causal Process		Desired Effect
Successful program:	Program	*set in motion* →	Causal Process	*led to* →	Desired Effect
Theory failure:	Program	→	Causal Process	→⊣	Desired Effect
Program failure:	Program	→⊣	Causal Process	*(would have led to)* - - - - - -▶	Desired Effect
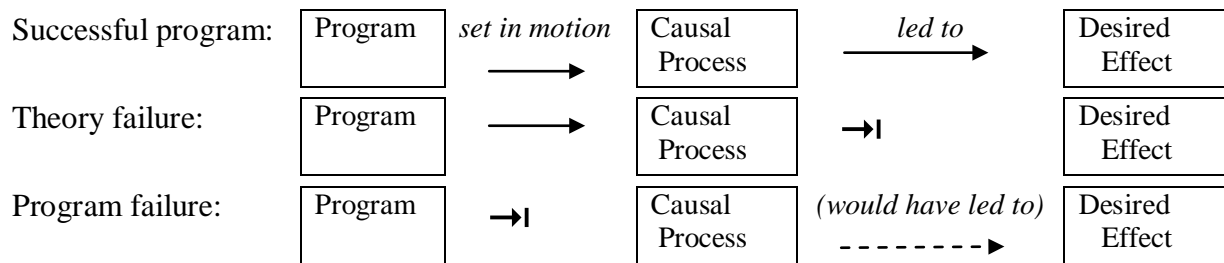
I distill this diagram into the shaded area between outputs and actual outcomes in my pyramid-shaped logic model.

Weiss notes that chronologically an evaluator might start with inputs, go to program implementation and interim progress markers, and proceed to desired and unintended outcomes. This sequence conforms both to Astin's I-E-O model and to Shadish, Cook, and Leviton's (1991) breakdown of programs into three components: 1) internal program structure, 2) external forces that shape programs, and 3) understanding how programs change to enhance social goals (Fitzpatrick, Sanders & Worthen, 2004, p. 54).

Weiss alludes to generalizability of inputs when she says, "Inputs can serve meta-analysis [as] aggregate results of evaluations of similar programs" (p. 133). One premise of this dissertation is that practices and models can be borrowed and customized (although not universally generalized). Patton (1990) joins in progressing from generalizations to extrapolations: "Generalizations decay…, valid only as history", and always context free (Patton, 1990, p. 486, citing Cronbach 1975; Guba & Lincoln, 1981, p. 62). Crediting Cronbach, et al (1980) for this insight, Patton elaborates that effective evaluation should strike "a middle ground stance to permit extrapolation, which connotes that one has gone beyond the narrow confines of data to think about other applications and findings" (p. 489).

43

With an infinite permutation of shapes and sizes, logic models not surprisingly may be categorized by approach (WKKF, 2004, pp. 9-10):

1. Theory Approach Models emphasize the theory of change that has influenced the design and plan for the program…. These models illustrate how and why you think your program will work. They are built from the "big picture" kinds of thoughts and ideas that went into conceptualizing your program. They are coming to be most often used to make the case in grant proposals. Models describing the beginnings of a program in detail are most useful during program planning and design.

2. Outcomes Approach Models focus on the early aspects of program planning and attempt to connect the resources and/or activities with the desired results in a workable program. These models often subdivide outcomes and impact over time to describe short-term (1 to 3 years), long-term (4 to 6 years), and impact (7 to 10 years) that may result from a given set of activities…. [These models] are most useful in designing effective evaluation and reporting strategies.

3. Activities Approach Models pay the most attention to the specifics of the implementation process…. These models describe what a program intends to do and as such are most useful for the purposes of program monitoring and management…. Models that emphasize a program's planned work are most often used to inform management planning activities... [and] program development.

Kellogg Foundation authors find commonality within these approaches in explanations of underlying program assumptions. The model I employ in this study is best classified as an outcomes approach model. Note also a relative complexity of logic models in terms of layers and levels, not only of elements or ingredients. Astin's I-E-O arrow represents assumptions, among many factors that link inputs and environment to intended outcomes. No longer lumped into a catch-all category of variables and prompts, those same assumptions can be depicted readily and separately in elaborate logic modeling. WKKF (2004) affirms that assumptions correspond to "big ideas" as opposed to "nuts and bolts" (p. 10). Consequent logic models might sprawl across multiple white boards in a university planning office's strategy room, but – as I will demonstrate – logic models are presentable as frameworks that inspire narrative.

In contrast to the depth offered by logic modeling, Astin's model does not discern outputs and outcomes. Rather, WKKF (2004) distinguishes outputs as measurable units of information – numbers of participants, dollars earned – and outcomes as changes in attitudes and behavior (p. 8). With steady assessment to determine causality, outcomes might manifest further as impact.

In using program theory as analysis guide, Weiss (1997) recognizes a possible limitation that data are not thoroughly accommodating.

> Associations are usually partial; therefore exercise judgment in figuring out the extent to which they support the theory. Program theory has many benefits for the evaluation enterprise. Even when it does not lead to crisp conclusions about the processes and mechanisms of change, it will provide useful information, and encourage analysts to develop new methods to fit between theory and events (p. 278).

Program theory describes the logical progression toward expected effects. The framework consists of measures of inputs and outputs, appropriate measures of intervening processes, good statistical know-how, and opportunity to add on further inquiry when results are puzzling (Weiss, 1997, p. 290).

Program theory and logic modeling may be especially helpful in addressing "the nature of assumptions that underlie programs, an iterative process that rethinks priorities and choices" (Weiss, 1997, p. 323). Until I encountered logic modeling, I had pictured linear evaluative processes that culminate with outcomes, as in Astin's I-E-O model. Logic modeling allows outcomes to fold and reflect back into criteria based on sound and malleable assumptions. A linear model corresponds to a paper map from "Triple-A", while logic modeling is the Google "street level" map. Both utilities enable the traveler to reach a destination, but the logic model gives a truer glimpse in advance.

This different paradigm alleviated my concerns about outcomes-only assessment and fostered my appreciation for planning that began with vision of long-range goals rather than hope

and expectation that programming would result in favorable outcomes organically. This conception permits mission and vision, not process, to drive programming in Student Affairs or perhaps other collegiate contexts. A logic model can portray how mission, vision, and values drive the assessment cycle for the university or unit (Patton, 1990; Weiss, 1997; Frechtling, 2007).

Perhaps Patton (1990) gives the most terse and accurate description of logic modeling as a "summative test" (p. 107). In this definition, Patton refers to the logic model's direct and simple ability to diagram what is working, and not working. I might supplement that definition with observation that elements can fit into categorical lists representing formative enhancements and repairs to a healthy campaign that has passed summative muster. In achieving both formative and summative purposes, logic modeling can be an effective means of meta-evaluation. Patton offers reminder that "evaluation syntheses are not an end in themselves but a means of generating powerful insights about effective program practices and processes across multiple experiences and cases" (p. 428, italics removed).

➔   In summary, logic modeling can accomplish both practical and research-based benefits (Frechtling, 2007, p. 6). Logic modeling serves that dual purpose in my study. First, it is the heuristic I employ to meta-evaluate the assessment campaign described in the case study. Second, my results might support recommendation to use this or another logic model for the Student Affairs department's ongoing assessment phases.

Up next, the Methodology chapter offers rationale and maps out how I merge a case study and an original logic model into a cohesive analysis of how thoroughly the profiled Team conducted its evaluations of outcomes in accordance with the Student Affairs department's mission and vision.

3.0   METHODOLOGY

Weiss (1997) calls methods "bundles of techniques that can be put together in different combinations; not necessarily design packages, total entities to be selected" (p. 180). As many various designs might work for an evaluation project, different designs or combinations of methods might structure and inform an assessment campaign. This sub-chapter covers rationale and origin of one particular case study: why and how a department has chosen to invest in expensive and labor intensive assessment activities, among numerous opportunity costs and alternatives. Specifically, this chapter describes how I plan to fuse an original logic model and case study into new and interesting research on student affairs program assessment. Subsequent analysis will demonstrate how logic modeling facilitates meta-evaluation of accomplishments featured in the case study.

3.1   CASE STUDY METHODOLOGY

By presenting expert testimony on why case study is good, this sub-chapter justifies why I have chosen case study as an effective way to test how logic modeling can inform and possibly guide a complete assessment campaign.

Maria Piantanida and Noreen Garman (1999) observe that a researcher heeds the call of one research paradigm or genre over another. My choice of methodology translates stated

purposes and goals of evaluation into practical application in student affairs – and ultimately raises awareness, acceptance, and expertise in these techniques as integral components of university planning. In an earlier, less informed stage of my research, what appealed about case study was its flexibility to embrace more than one domain. One could describe a case that illustrates certain evaluation principles. Another researcher could use different or same case(s) to demonstrate student affairs practices. Most economically a researcher can use the same case(s) to cross-tabulate variables over both domains. Case study structure and approach not only build knowledge but also allow sharing and comparing findings – which are seldom universal or interchangeable among contexts. Robert Stake (1978) justifies case study in that "… people have arrived at their understandings mostly through direct and vicarious experience" (p. 6). As Patton (1990) argues, "Case studies are particularly valuable when the evaluation aims to capture individual differences or unique variations from one program setting to another, or from one program experience to another" (p. 54), I would argue the flip side, that case study can illustrate commonalties leading to practical standards and consensus that bolster meta-analysis and theory.

Although many different qualitative methods could capture direct and vicarious experience, case study seems apt in depicting processes familiar to my intended audience of evaluation practitioners and theorists who seek efficient and timely ways to achieve specific outcomes. This familiarity may explain why "direct and vicarious experience" carries value. Practitioners may feel too busy to test all the same waters, and instead may learn from others' experiences. Patton (1990) claims that a goal of case study is to "assimilate information from various sources to produce a highly readable narrative that could be used by decision makers and information users to better understand what it was like to be in the program. Different sources

48

cross-validate findings, patterns and conclusions" (p. 389). I watched this type of collaboration and cross-checking occur in Team activity.

Stake (1990) distinguishes between three types of case studies: intrinsic, instrumental, and collective. If I were confident that a particular student affairs unit demonstrates exemplary and possibly universal policy, then I would take an intrinsic approach, which illustrates the merits of one case and allows a self-contained story to unfold[19]. However, intrinsic methodology does not apply here, because my chosen context may not warrant singular attention. One university's brand of student affairs evaluation may or may not have meaning or application to other institutions.

Why, then, not a collective study? Rather than focusing on one example, I could survey or interview representatives from a broad range of American universities, and thereby arrive upon some measure of consensus. The likely result would be a practitioners' matrix or checklist. However, a study that broad could dilute one contributor's story. Without embarking on a fully collective approach, I may discover in future research that additional steps and peer comparisons accomplish supplemental triangulation for a collective study.

My study fits best into the category of instrumental case study, because some practices might translate well into contexts beyond my primary focus, while others may not. Stake (1994) says, "…The case assumes a place in the company of previously known cases" (p. 443). One aim of the instrumental version is to investigate and perhaps suggest generalized, repeatable guidelines for wider application – with the realization that case observations are seldom if ever universal when applied to unique, peculiar contexts. To identify generalizable features, the

---

[19] Examples of good intrinsic studies might include Salk's discoveries of a polio vaccine, Fossey's account of "gorillas in the mist", or Duneier's *Sidewalk* (2000), a look at New York City homeless culture. A study with intrinsic value is one that a fellow researcher or even a layman might want to read for its own aesthetics or statement about the human (or primate) condition. Standing alone, the intrinsic study adds to discourse.

researcher must carefully heed "a sampling problem. The case will be selected to represent some population of cases…. The cases are opportunities to study the phenomena" (Stake 1994, p. 446). An instrumental approach strikes me as a manageable and interesting compromise between the confined context of intrinsic case study and the breadth of a collective approach.

Another appeal of the instrumental approach is, as former co-chair of an active assessment committee, direct access to meaningful data and processes. Here, data consist of official reports, support documents, meeting minutes, and e-mail correspondence. Data and intent drive the process, rather than a researcher attempting to fit piecemeal data into patterns typical of a favored methodology. In other words, I did not elect case study because of actual preference or facility, but because analysis of a student affairs department's interlocked activities required a comprehensive and cross-disciplinary approach. My incentive was not convenience, but immersion.

Stake (1994) describes a structured tradition in case study, including a six-part "recipe" of conceptual responsibilities (p. 448):

1. Bound the case;
2. Identify prominent themes and issues;
3. Seek data patterns;
4. Triangulate;
5. Investigate alternative interpretations; and
6. Develop assertions and generalizations.

Heeding Patton (1990) and other theorists' advice, I would substitute the word "extrapolations" for generalizations. Stake adds what might seem an intuitive note: "My choice would be to take that case from which we feel we can learn the most" (p. 446). Regardless of its reach and duration, a thorough and thoughtful case study is characterized by its reflective or

interpretive nature. Stake (1994) advises "pondering the impressions, deliberating recollections and records" to derive deeper meaning than can be obtained by simple transcription.

➔   Through narrative case study, I hope that this story will share useful shortcuts, add to ongoing discourse among student affairs and evaluation professionals and scholars, and offer a flexible and versatile framework to guide other assessment campaigns.

3.1.1   **Bounding the case**

This sub-chapter distinguishes environment from periphery, to allow focused analysis. Without logical and measurable fit, peripheral elements lie beyond scope of the environmental components of program theory espoused by Weiss and Patton. An illustration is where a community college might be tempted to borrow assessment practices from Harvard or Michigan. Such an assessment philosophy may ignore the logical bounds of environmental context. While certain principles (i.e., "conduct a good study" (Upcraft & Schuh, 1996)) may translate from one venue to another, over-emphasis on peripheral wisdom may compromise or capsize an otherwise successful assessment campaign. An evaluator who has conducted a needs audit based on extrapolation may diagnose the incorrect ills and subscribe impractical solutions. Outright generalization should surely face summative boot. Instead, let "local conditions prevail" (Bracey, 1997, p. 133).

The investigator has no choice but to start somewhere, and allow variables, interactions, and permutations to generate new tests and subsequent findings. I bound this case study by presenting the viewpoint of a participant-observer, as co-chair of a formal Assessment & Measurement Team (hereafter, "the Team") from August 2005 to June 2007 – two complete

seasons of an ongoing assessment cycle. I think that my immersion and actual stake in the profiled assessment campaign contributes richness beyond a usual, perhaps journalistic narrative.

Within Student Affairs departments[20], independent evaluations took place on a regular basis. While these projects were part of an overall picture of university department programmatic and academic assessment contributing to stated mission and vision, **this study focuses on the Team activity centered around Student Baseline Outcomes**. The diagram below illustrates how unit and team level evaluations fit together, using Residence Life (my professional base) as an example.

In the figure below, ***the green box represents the scope of this study***; every other element in this diagram represents a related but separate topic (perhaps for future research or a textbook chapter) tied to the department's official mission toward a "culture of assessment".

Figure 9   **Residence Life evaluations as subset of Student Affairs assessment process**

---
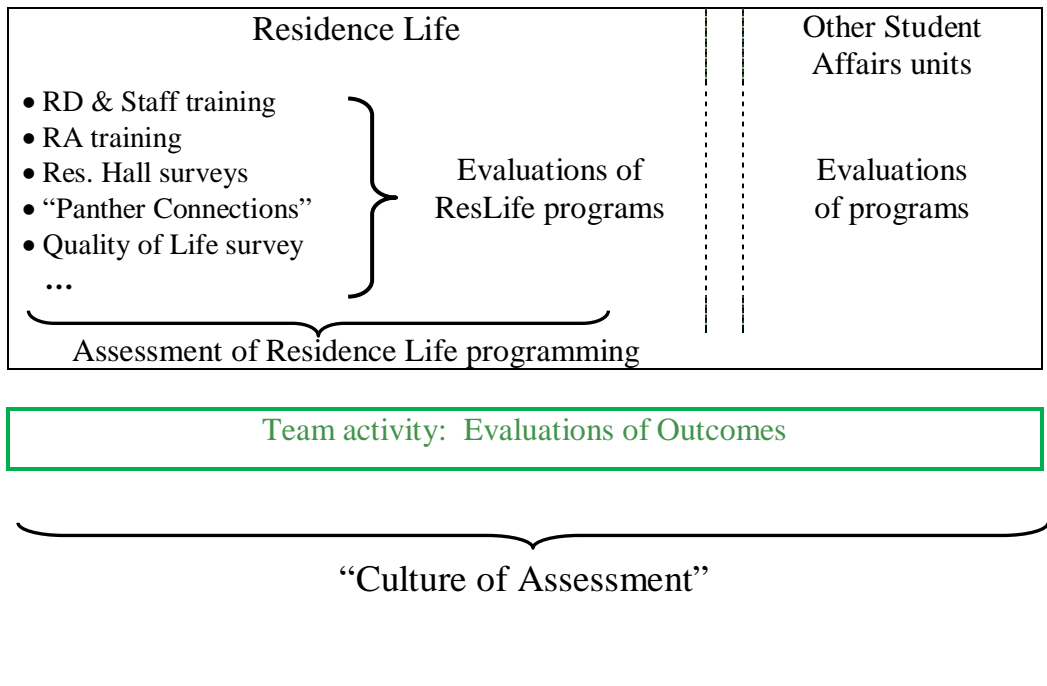
[20] Student Affairs consists of Student Life, Residence Life, Career Services, Student Counseling, the International Office, Disability Services (DRS), Student Health Services, Judicial Services (USJS, absorbed into Student Life as of summer 2006), and Office of Cross-Cultural and Civic Leadership ("O3CL", established October 2006). Each Student Affairs unit is led by a Director who reports to the Dean.

3.2   BUILDING AND APPLYING A LOGIC MODEL

An underlying premise of this study is that student affairs practitioners and planners can identify useful and appropriate evaluation and assessment practices that enhance effectiveness of programming. This research explores the extent to which evaluation and assessment principles have applied to a specific university department. Logic modeling is but one tool that allows planners and evaluators to present and organize effective ways to accomplish assessment goals. A good model can encourage the evaluator to seek beyond convenient solutions and bound the case into realistic time frames and targets.

Applying one particular logic model, I investigate how Pitt Student Affairs has pursued assessment objectives, follows Stake's heuristic, and may represent a comprehensive and worthy assessment campaign. As a component of meta-analytical dialogue, analysis of the Pitt case thereby lends insight on how planners might implement or fine-tune ongoing assessment in Student Affairs and perhaps extend favorable, tested practices into related strategic units.

Analysis attempts to fit everything that happened in this case into appropriate layers, connections, and labels according to the logic model below. I show that this model can be used to analyze specific evaluations and also combine findings into a meta-evaluation of the whole campaign. The figure below restates the model I use to accomplish this fit.

Figure 10   **Logic model depiction of assessment cycle**
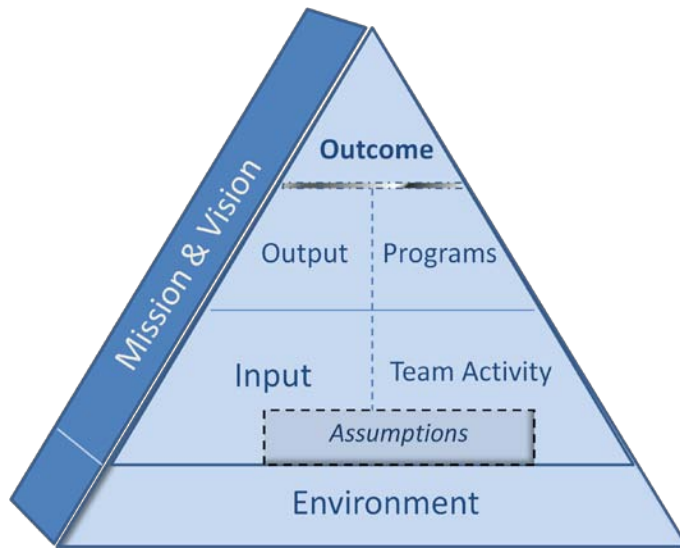
As the "spine" of assessment process, mission and vision drive activity and frame outcomes, within environmental constraints. In this case, each and every Student Baseline Outcome may be prefaced by the official statement of vision "to provide University of Pittsburgh students with the best collegiate experience in the country."

Note that this logic model contains all three elements of Astin's model. However, my intents are to distinguish outcomes from outputs and to fashion layers and directionality, resulting in more durable analysis. The logic model is configured with "environment" as the base layer. I feel that an external scan is a sensible place to start any evaluation; if an evaluation team is lost for ideas, looking at environmental factors can be a default activity. The discovery process will probably generate useful knowledge that triggers investigation of other layers and elements. Underlying assumptions about inputs, outputs, and favorable courses of interrelated activities determine program formulation and implementation. The "environment" layer includes accepted standards for evaluation. While not gospel, standards are compelling environmental factors, as are conventional wisdom and trend-setting by peer and aspirant institutes. The "environment"

box incorporates political factors that frame but must not dominate evaluation processes. The "boundaries" of logic model components are translucent to allow practitioners the latitude not to shoehorn components into a category. Rather, locations of inputs and environmental factors and the outcomes they promote can be situated logically.

This logic model offers a high-level view of process and linkages flowing upward, culminating in achievement or status of short-, medium-, and long-range outcomes. The shaded area between outputs and outcomes represents causality, which must be proven experimentally, presumed, or left to speculation lest outcomes be rendered "unintended" (Patton, 1990; Weiss, 1997; House & Howe, 1999; Upcraft & Schuh, 1996). Impact is represented by a trickle-down effect of outcomes back upon reconfigured assumptions and follow-up evaluations, as shown in the figure below:

Figure 11   **Logic model depiction of "impact"**
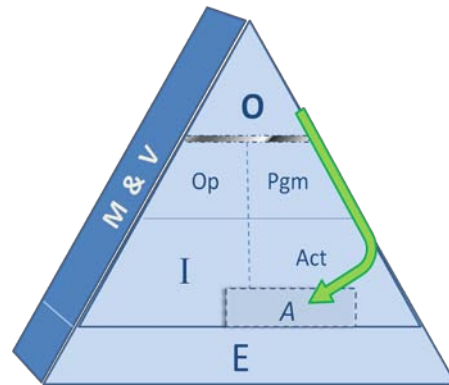
The bulk of analysis fits into the "program" and "activity" slots, with a wealth of documents and institutional knowledge as the primary source of data[21]. Such activity includes

---

[21] Patton (1990) points out that for any level of observation, field notes are essential: "Don't leave it to recall…. It is critical not to be deceived into thinking that those feelings can be conjured up again simply by reading the descriptions of what took place" (pp. 239-241). It is unfortunate, in retrospect, that I did not consider the value of more thorough field notes at the onset of Assessment Team meetings, as I had not yet taken a class in field methods nor read Patton's book. Therefore, minutes must suffice for descriptions of meetings and Team dynamics, while a nostalgic visit back to our conference room conjured up images of the setting. The lesson I draw here is that no research is perfectly planned, but the researcher can resort to next best sources and data gathering techniques.

interviews and results, surveys, internal and external scans, and formal reports. "Outputs" include the documents themselves, reactions to recommendations revealed in follow up minutes and memos, press releases on policy, actual policy change, and program enhancements.

Data are available from artifacts collected for the first two seasons of the assessment campaign, from August 2005 to June 2007. Supporting documents, official reports, meeting minutes, and transcripts comprise the data sources for this case[22]. I supplement this data with personal reflections. According to Patton (1990), such records and documents represent a rich source of information, "a trail of paper that the evaluator can follow and use to increase knowledge and understanding about the program" (p. 233). Comparing actual results to written directives enables an evaluator "to point out what did not happen when program goals, implementation designs, and proposals suggest that certain things ought to happen" (p. 235, emphasis removed). That is, the evaluator's judgment and experience can alert an absence of an essential activity or factor (p. 236). Weiss (1997) presents another efficiency argument: Using existing data avoids recruiting and maintaining comparison groups made of real people (p. 207), at considerable opportunity cost.

As records and documents are the primary data source for this case study, then the data from documents, correspondence, and recollections that occupy slots in the logic model flesh into lengthier narrative that interlocks with analysis of other outcomes. Such narrative comprises this study's "Analysis" chapter. The Student Affairs office depicted in this case study defined 14 outcomes. Hence, complete analysis could consist of 14 sub-chapters. Brevity and economy instead call for a representative sample of outcomes: I distill analysis into a manageable three sub-chapter discussion of the Team's first evaluation effort, of an unsuccessful attempt, and of a

---

[22] Signed, official access to these artifacts has been granted by the Dean, the "keeper" of this data. A copy of that permission document was submitted for IRB approval, granted in June 2008.

strong evaluation. Appendix B compiles the complete set of all 14 evaluations that derive a cumulative "scorecard" ahead in Table 7. In the last step of meta-evaluating Team performance on individual outcomes, I rate quality and thoroughness by using the metric of symbols shown below.

Table 3   **Key to logic model "scores" in meta-evaluation**

Symbol	*Meaning*
✓	Element has been analyzed adequately.
✓+	Team did an exemplary job in this task
( )	Element is unresolved or conditional.
-	Element or task was not resolved.

The logic model gets filled in with symbols denoting qualitative judgments. The resulting model may look like the example figure below.

Figure 12   **Example meta-evaluation of an evaluation process on one outcome**
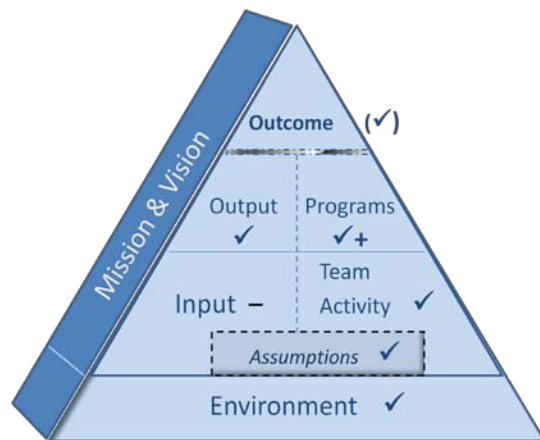
In this example, the Team used sound assumptions (check), conducted or cited a sufficient external scan (check), located an exemplary wealth of programs (check plus!) related to this outcome, offered a report as output (check), and found preliminary basis for the outcome to be achieved (task completed, but result pending; I use symbols in combination, where appropriate). That is, the parentheses indicate that the outcome has not actually been achieved

yet, although the Team has sought evidence on progress. The minus sign in the "input" slot tells the project leader that a lapse in this process involves student needs and characteristics. More detailed narrative would explain what was missing from the Team's effort on this evaluation.

The final piece of analysis is to meta-evaluate the whole campaign. I accomplish this task by tallying the series of logic model ratings into a matrix (Table 7, p. 120) that summarizes overall meta-evaluation of performance. Totals tell both how the Team performed per outcome and per element. By reading across rows, the researcher sees how the Team scored on an outcome. By scanning down columns, the researcher sees how the Team performed on a particular analysis element or component. For example, checkmarks in all or most "environmental" slots indicate that the Team demonstrated prowess with external scans, while several minus signs would indicate that the Team could have been more thorough with this task.

The process of interwoven lists strikes me as potentially neat and orderly, as long as I heed Patton's (1990) warnings that a "logical system will create a new typology whose parts all may or may not actually be represented in the data.... It is easy for a matrix [or diagram] to begin to manipulate the data as the analyst is tempted to force the data into categories created by cross-classification to fill out the matrix [or diagram] and make it work" (pp. 411-414). This is a trap the evaluator must avoid, as not to generate misleading conclusions by improper and convenient categorizations. Patton also warns that an "empty cell can alert or sensitize the analyst to the possibility of activity or behavior overlooked…, or logically possible but not manifested" (p. 414). Empty cells and null values are not congruent in this model – more likely a prompting to triangulate sources or dig deeper. Patton suggests using matrices for organizing outcomes not clearly articulated or uncertain to unfold (p. 416), rather than listing identifiable ones.

The resulting series of models and accompanying narrative resolve questions about the assessment campaign such as, "Was the process complete?" "What gaps must be filled to achieve desired outcomes?" "Do these results indeed serve mission and vision?" From these elements should emerge a cohesive story, with evidence to support or challenge my hypotheses on the logic model's utility to analyze and meta-evaluate a student affairs assessment campaign.

➔   In sum, my premise is that retro-fitting elements of the case study into a series of logic models and related narratives achieves my research objective of meta-evaluating the campaign presented in this case study. Using this logic model heuristic, analysis weighs what the university system depicted in case study has accomplished. This logic model "roadmap" may also speculatively guide the profiled system's future phases of assessment and possible success in achieving desired outcomes still pending or due for fresh evaluations. In the spirit of meta-analysis, my hope is that readers find descriptions and lessons to compare and contrast to their own evaluative practices, adding to knowledge and consensus about current process and program theory for university student affairs assessment campaigns.

Patton (1990) summarizes how a logic model can accomplish the task. The guiding principle in using a logic model is to recognize that "the linkage between processes and outcomes is a fundamental issue in many program evaluations" (p. 415). He continues,

> Conceptualizing program outcomes and impacts can also be either an inductive or logical process. Inductively, the evaluator looks for changes in participants, program ideology, distinctions among participants over 'those who are getting IT' and those who aren't, where IT is the desired outcome. Such outcomes as 'personal growth', 'increased awareness', and 'insight into self' are difficult to operationalize and standardize (p. 420).

Enlisting a fresh logic model, now I proceed to show how one university student affairs unit got IT done.

4.0   ANALYSIS

This study segues into its story portion: how I, as participant evaluator, oversaw a new assessment campaign and how I reflect upon its accomplishments. A distinguishing feature of this study is the insider's view of the process and actual stake – a sense of shared ownership and belief in the project's merits.

In this chapter I apply the logic model heuristic described in the Methodology chapter. Sub-chapter 4.1 presents the evidence that drives analysis – the facts of the case. Next, sub-chapter 4.2 synthesizes the case study and logic model, fusing facts to theory; examining links between inputs, activities, and outputs; and driving conclusions and recommendations.

4.1   THE CASE

This chapter fleshes out case details and the timeline glimpsed earlier in the "Purpose / context" sub-chapter 1.1.2.

The University of Pittsburgh ("Pitt"), a public and privately funded urban campus and research university in the mid-Atlantic region[23], may have some good ideas how to ingrain

---

[23] In keeping with a bounded approach advocated by Stake (1994), this case study looks at only the main campus in Pittsburgh's Oakland neighborhood, and not at satellite campuses in Greensburg, Johnstown, Bradford, and Titusville, PA.

evaluation and assessment into back-office culture and operations. This case involves Pitt's Student Affairs department's first two seasons of ongoing evaluations of programs, from August 2005 to June 2007, when I served as co-chair for the Assessment & Measurement Team (hereafter, "Team"). This study warrants brief description of the physical and administrative structures of Pitt Student Affairs.

In addition to the showpiece Cathedral of Learning, a convention and athletic "events" center, the Student Union, the medical school complex, graduate schools, various classroom buildings and open spaces, the Pitt campus consists of twelve residence halls[24], in which dwell the 5,250 on-campus undergraduates (according to Residence Life hall rosters, 2006-08). Resident Directors (RD's) oversee each hall, or clusters of halls. Each RD supervises a team of Resident Assistants (RA's), undergraduate students who apply and interview for these competitive, high-profile positions. RA's are expected to take charge of communications, teambuilding, community spirit and morale, and formal programming on their "floors".

Student Life and Residence Life units add formal structure to in-hall programming efforts. Student Life events are generally campus wide, such as freshman orientation and homecoming. Student Life includes the fraternity and sorority councils, treated in this study as an off-campus branch (that is, not technically part of "input" factors but interested stakeholders). Residence Life meanwhile oversees specific, customized, impromptu programs. Residence Life often needs only to provide funding for refreshments and supplies (such as movie and video game rentals, paint, pumpkins, ingredients for baked goods, etc.), arrange room reservations, and design scripts and publicity materials.

---

[24] Student Affairs has been vigilant not to call these communities "dormitories".

In addition to Residence Life and Student Life, Student Affairs consists of Career Services, Student Counseling, the International Office (OIS), Disability Services (DRS), Student Health Services, Judicial Services (USJS, absorbed into Student Life as of summer 2006), and Office of Cross-Cultural and Civic Leadership ("O3CL", established October 2006). Each Student Affairs unit is led by a Director who reports to the Dean. The staff roster lists 170 names, including nine Directors, as of July 2008. These tallies have not changed much since 2004. Unlike many peer and aspirant institutes, Pitt maintains a separate and autonomous Housing division – which often collaborates with Student Affairs on projects such as an annual Quality of Life survey that includes feedback on dining services and hall maintenance. A major challenge for Student Affairs and Housing is to keep administrative distinctions transparent for students – who do not care which unit is in charge of ensuring quality meal plans and changing light bulbs, but simply that it gets done.

Before arrival of an energetic new Dean, Student Affairs had enlisted modest measures of evaluation to ensure quality programming. According to testimony, these measures consisted mainly of informal and observational monitoring. The new Dean raised the stakes, envisioning a "culture of assessment". That recurring buzz phrase in early speeches and all-staff meetings became a calling card, a unifying mantra for Pitt Student Affairs as an official assessment campaign was launched.

4.1.1 **Assessment Campaign**

One assumption is that any university administration sees planning and evaluation as potentially favorable, useful practices – but not as core-critical as curricular development and faculty recruitment. A factor that might keep regular assessment from becoming "standard practice" is

the necessary distraction from usual administrative burdens that are more urgent and visible. If planners glimpse actual reward on a distant horizon, then short term objectives may require administrators to re-train towards evaluative practice. That costs time, money, and effort.

Pitt Student Affairs decided it could afford the investment. To serve efforts that might actualize the envisioned "culture of assessment", the Dean designated several new teams[25] during the summer of 2005. Each of these five new ad hoc committees consisted of up to twelve representatives from all units of Student Affairs. Positions were voluntary and non-salaried. This arrangement saved the department costs, in comparison to creating new salaried positions on permanent task forces or enlisting pricey outsourced consultancies. Incentive for members was often justified as training and development. Once designated, teams enjoyed considerable autonomy in scheduling meetings, setting agendas, and establishing project goals and timelines. All teams included liaisons, who were experienced upper-level managers or directors and who would communicate concerns and progress to decision makers. Allowing staff members to select which team(s) they would join resulted in best strengths and satisfaction among participants and, in theory at least, maximized productivity for all teams.

Appointed co-chair of the Assessment and Measurement Team[26], I found myself in position to gather and interpret data from various systems. While I oversaw research conducted by Team members, I was accountable for overall monitoring, implementation, and reporting. Program structure, funding, personnel, politics, and communication channels determined sampling and data collection – for example, access to confidential data and documentation. The

---

[25] I distinguish teams from task forces or working groups, which formulate measurable outcomes in line with institutional mission. The teams meanwhile focus on process to measure those outcomes. The other teams were Staff Recognition, Training and Development, Marketing and Communications, and Information and Technology.

[26] Upon reflection I wondered how I, as a graduate student among seasoned veterans, became chair. Evaluation had already become my specialty, after a year of projects in Residence Life. Evaluation was meanwhile still a mysterious concept and process to most employees. Although I volunteered, my candidacy was accepted immediately by my Director and resoundingly by other Directors. It was a logical career move for me.

Dean's vision was that this Team's findings could have major impact in justifying funding and resources toward various programs in Student Affairs. In terms of level of influence, we operated as an advisory council rather than being granted executive privilege over assessment matters. While implicit for the first two years, the Team's advisory status was made formal in the third year of the assessment campaign, after my departure.

Elevating evaluation and assessment as strategic directives was intended to promote university vision. Here, Student Affairs strove "to provide [u]niversity… students with the best collegiate experience in the country" (Vision Statement, 2005, p. 1). Evaluation could measure the success of that bold vision in qualitative and quantitative terms, including peer comparison.

While the Team had energy, one could inquire whether we had experience or a plan. Before applying the logic model retroactively and proactively to the Team's activity, as I will do next sub-chapter, it may be helpful to determine whether the Team followed any rudimentary plan at all. Weiss (1997) offers a list of essential conduct or activities (p. 72):

1) Identify key questions of study;
2) Decide upon quantitative, qualitative, or combination design;
3) Develop measures and techniques to answer research questions;
4) Figure how to collect data to operationalize measures;
5) Create research design, including comparisons and timing;
6) Conduct actual collections and analysis;
7) Write and disseminate a report;
8) Promote appropriate use of results.

A sensible beginning to our practice might have been a deliberate needs assessment or assessment audit[27] (Lane, 1998). Our needs audit had already been implicitly conducted in the Dean's creation of the Team. Remarks to the Team at our opening meeting pointed in the direction of a further, deeper audit, framed by "Student Baseline Outcomes" compiled during the

---

[27] I prefer the contraction of those two phrases into "needs audit", as not to create a circular definition of assessment.

summer of 2005. Our goal would be to evaluate each outcome by whatever means we saw fit. This directive indicated that much of our early activity would be formative. The evaluator or team could initiate one of two approaches to initial data gathering:

Cooper's Evaluative Assumption #1: In absence of evidence, never assume that administration has adequately and accurately analyzed (i.e., via comprehensive internal scan) and included all input and environmental variables from which logical outcomes flow. Challenge assumptions. This assumption corresponds to a summative approach.

Cooper's Evaluative Assumption #2: If outcomes seem prima facie realistic, assume that administration has laid necessary groundwork from inputs and environmental factors. Trust assumptions. This assumption drives a formative approach – at least initially.

The approach may depend on the amount of time allotted, expenses allocated, and other practical considerations. In ideal scenarios, Assumption #1 should dictate that the evaluator conducts a comprehensive internal scan, or locates evidence of a comprehensive scan by task forces or Student Affairs decision makers. Attentive documentation by project planners should have that information readily available.

However, practical circumstances and management philosophies may dictate that an evaluator accepts well-formulated outcomes as a starting point for analysis and performance measurement. Should an evaluator risk offending his client by suggesting that input and environmental factors were not sufficiently analyzed or consulted? That answer depends on the evaluator's standing in the department. If that culture deems the evaluation process intrusive, then the evaluator is wise to take a non-confrontational stand. Then the task is to gather evidence that suggests formative validation of or improvements to sound policy.

The Team began evaluations with little more than these assumptions as basis. Given political circumstances and the newness of departmental assessment, we deferred to an "Assumption #2" formative approach. That is, we accepted outcomes as soundly formulated, then proceeded with steps to answer specific questions and gather supporting details. Assumption #2 earned a corollary: Where a needs audit – formal and organized, or based upon ongoing observation, documentation, and testimony – has already been conducted or presupposed by the time an assessment team is assembled, support teams' default assumption should be to trust directors / agents to have conducted this audit properly.

Before explaining in detail how our team reached consensus on approaches and priorities, it is helpful to describe team composition and dynamics.

4.1.1.1 Team dynamics

The Team's first meeting took place in early September 2005, at the beginning of the new academic term. The Dean's introductory statements revealed that the idea of an assessment team had spawned creation of all five ad hoc committees. Comments amounted to a pep talk. Dominant themes included the following:

- The team recognizes its primary task to observe and measure whether the division reaches specific, stated outcomes. Of these, "some will be accomplished; all will have progress." (Student Affairs meeting minutes, September 8, 2005, p.1). Another guide for benchmarking would be an Annual Planning Document scheduled for update in the following spring term.

- The team approach prevents personal ownership by departments and individuals. While Directors and the Dean would regularly offer insight on how to accomplish each task, the

process grants the Team autonomy and judgment to forge best assessment practices. The Team could start anywhere, with support of receptive and informed leadership that would boost the likelihood of subsequent buy-in.

- At the same time, leaders envisioned a primary internal outcome to promote a cohesive, unified division, instead of several units. The Team would lead in creation of a "culture of assessment" within our operations. Efforts and dynamics represent a mirror; "we cannot deny what we see". (Student Affairs meeting minutes, September 8, 2005, p. 1).

- Foremost, assessment would serve the Student Affairs vision "to provide University of Pittsburgh students with the best collegiate experience in the country" (Pitt Student Affairs, Statement of Vision, 2005). Earlier incarnations of the Student Affairs vision statement included "educat[ing] the whole student" based upon four key attributes: communication skills; sense of motivation; sense of responsibility; sense of self. While these touchstone concepts may have been rephrased, their emphasis did not change or diminish throughout the assessment campaign. (As the reader will see in sub-chapter 4.2.3, these attributes resurfaced in a reconfigured "Pathway" program.)

If the Team took our task seriously, these themes would carry through every successive meeting. The kickoff meeting proceeded with a lesson on basic assessment and evaluation principles, conducted by the two co-chairs, and centering around knowledge I had gathered from coursework on evaluation and assessment. After agreement on team members' roles – chairs to cover documentation, archiving, and communications; team members to become data gatherers and sharers of ideas and best practices; liaisons to serve as an advisory and, when necessary, advocacy resource (i.e., for additional funding) – the meeting proceeded to project proposals, current practices, and benchmarking. This impromptu internal scan pinpointed topics

foremost under discussion throughout Student Affairs (i.e., policy on underage alcohol violations and collaboration with Judicial Affairs), established consensus definition of "student affairs" – including intent for a clearer "brand name" identity – and shared information on unit-level evaluations and instruments. Concluding remarks focused on proper pacing to avoid feeling overwhelmed; a committee "cannot expect to cure all that ails Student Affairs, and should instead observe limits and seek to take carefully measured strides in positive directions" (Student Affairs meeting minutes, September 8, 2005, p. 2). The Team adjourned with "homework" for members to meet with their respective directors to discern active programs pertaining to achievement of division outcomes. Despite our broad ideals and naïve understanding of assessment cycles and process, I think in retrospect that this was a favorable start with a positive attitude.

One issue that arose was how to share classified documents. While evaluators might be used to reading between lines of secrecy, this department maximized efficiency and mutual trust by opening the vaults to an internal commission. This latitude might not be as easily granted to an outsourced evaluator without additional paperwork on confidentiality.

The process that evolved most comfortably was for the Team to meet on a scheduled pace. Consensus determined that meetings would be held once a month, on the first Tuesday, and sometimes more often as needed. This structured approach allowed members to set individual timetables and block out times on calendars. This structure implied there would be no excuse that unscheduled or unanticipated meetings had surprised members into missing attendance. Where other priorities conflicted, chairs were informed in advance on most occasions. The Team never had more than three members absent for a given meeting, although one team member disappeared without notice. Two members resigned and were replaced the

following year by other staff members. The moral is that attrition was nearly non-existent, and the team remained tight[28].

Team members debated along a pre-released agenda formed by the co-chairs, who shared presiding and note-taking tasks. I usually wrote minutes from both sets of notes, and often we included submissions and clarifications from other Team members. Co-chairs and liaisons would review minutes for accuracy and clarity before distribution to the whole Team and Dean. Often other teams would request a copy.

The content of these minutes usually consisted of task assignments and target dates. Our tone remained collaborative when deadlines drew close. That mellow approach may be one common characteristic, for better or worse, of any non-compensated and voluntary team. The approach also seems consistent with Weiss' (1997) observations on qualitative discovery as an organic process as data unfold, where operations depart from expectations, and as new program theories develop (p. 266).

Accommodating task and project preferences allowed some Team members to cross-train and to explore unfamiliar areas, others to demonstrate expertise in familiar areas, and all to participate in a democratic, communal process of delegation. Sometimes Team members drifted into predictable roles of research. For example, some of the counselors with strong interviewing skills gravitated towards organizing and conducting focus groups and approaching students for man- or woman-in-the-street interactions. Others with strong writing skills helped with proofreading and online research for external scans. Minutes reflect that the Team recognized and utilized strengths as a good way to boost Team efficiency and productivity.

---

[28] Last I heard, my co-chair continues to serve.

I recognized that much of my background in business and software systems design could resurface in designing conceptual frameworks for project planning (albeit shy of an actual logic model) and crunching numbers. Besides delegating tasks into subcommittees (i.e., a six-person sub-committee each to investigate e-portfolios and mentoring outcomes), my role embraced time schedules and targets, identification of alternatives, and assignment of prescribed methods. These included internal and external scans, bounding the case and context, data collection, and seeking precedent and advice where necessary. Compiled into meeting minutes for distribution to the Team, liaisons, and Dean, this evaluation "recipe" served as a modest model for all or most of our evaluations. (See Appendix C, which displays some of the reporting protocols in Student Affairs.) I was engaging in program theory before I knew it had a name or application beyond my business background.

4.1.1.2 Student Baseline Outcomes

An early decision was which of 14 "Student Baseline Outcomes" to address first. (See the Student Baseline Outcomes document in Appendix A.) Decisions were based on imminence, availability of data that might inspire activity and momentum, and likely transition to related outcomes. A task force's formulation of outcomes and the Dean's official declaration of these outcomes had already satisfied a needs audit and set expectations. The outcomes document was the equivalent of our recipe book, or set of marching orders – although by Patton and Weiss' definitions these directives did not qualify as a framework. However, clear and assertive direction made evaluation seem straightforward. Priorities would likely change, but both parties had in writing what would be expected. Forthcoming analysis of outcomes will illustrate that each outcome has been deemed integral to student affairs strategy at Pitt. However, this quality does not guarantee that those outcomes have actually been fulfilled, a

judgment reserved for evaluation[29]. I think the list of outcomes passed a criterion posed by Popham (2008), that "aims must be incontestably important" because of limited resources (p. 133).

One deliverable intended from the onset was evaluation of programs cited in the "annual planning document", a collection of comparisons and benchmarking due in March of the following calendar year. Review of de-classified excerpts indicated close parallels between planning document directives and Student Baseline Outcomes. A consistency between various statements of mission and vision was a good sign. Looking back, this quality fulfilled a foremost requirement of a logic model, that consistency between source documents should reflect interlocked evaluation processes and separation of powers. The Assessment and Measurement Team would conduct evaluation projects that would give decision-makers necessary information to implement policy, rather than the Team being entrusted with also drafting a planning document from stated and measured outcomes.

I observed how a charismatic and confident leader can boost the expectation of success, in that evaluators had reason to believe that such outcomes have been well formulated. An expensive alternative might be to hire outsourced experts to customize packaged outcomes. Although the Team discussed philosophical reservations about outcomes-based assessment, stated outcomes allowed teams to unify around clear directives.

---

[29] Another worthy debate is whether 14 is a workable number of outcomes. Popham (2008) advises against "too many curricular targets" in favor of a "modest number" (pp. 128, 133). Twenty outcomes may have seemed an onerous task, while six would have seemed scant. Without delving into literature on optimal numbers of targeted outcomes for various contexts, I say that 14 seemed reasonable for this campaign. (Perhaps not coincidentally, for an accreditation project in my current job, our institution had to fulfill 14 objectives outlined by the Middle States Commission on Higher Education – credible triangulation for that number of intended outcomes.)

4.1.1.3 Team activity

Team assignments were external scans of peer and aspiration institutes and internal scans of our own practices. Part of the logic of representation from each unit was that someone could report on process in every unit and have that information collected into a single location and documentation. Reports of current practices could represent baselines for benchmarking purposes. One early internal scan involved locating useful instruments and then attempting to formulate new tools and usage surveys into standardized, customizable templates stored in a common-access database or "forms" tab on a website for easy access and review. Minutes from October 12, 2005, say that chairs and liaisons collaborated to collect instruments and templates into an archive. Submissions revealed how different and similar were various instruments and techniques used in Student Affairs units. Document review involved an extra session, between monthly meetings, for chairs and liaisons to review and categorize the wealth of data that Team members had gathered. Many of these early documents represented drafts, incipient attempts to fit existing practices into a likeliest outcome. Some documents showed a rudimentary understanding of evaluation that developed over subsequent formulations.

These activities contributed to collective evaluation strategies and a legacy of documentation that retro-fit into the "activity" layer of my logic model pyramid. Had I known at that time about logic modeling, bullet-pointed arrays of ideas could have taken on a logical, organized, and interactive format.

Next, Team discussions sought agreement on priorities and set target dates according to perceived urgency. Chairs and liaisons centered our third meeting (at the beginning of November 2005) around focusing our first evaluations upon e-portfolios and mentoring communities. We divided the Team into two subcommittees, taking input from members on

their preferences. Throughout the whole assessment campaign, this process would become a recurring and favored approach to multitasked directives: for Team members to select projects based upon the list of Baseline Outcomes and break into sub-committees who "owned" that task. We occupied our time purposefully, with less small talk than at previous gatherings, and delegated tasks by a planned, organized approach (i.e., agendas, cross-checking of details, and calls to resident experts on particular topics). We staggered tasks and time points rather than letting plans accumulate into an overwhelming pile. Although I think that the chairs' task management and delegation enabled the assessment process to run smoothly, a logic model could have mapped out the whole campaign, including target dates, instead of the step-by-step checklist approach that evolved[30].

Having scarce internal precedent for new programs envisioned for mentoring and e-portfolios, our attention shifted to external scans that would inform decision makers of compelling models, alternatives, sources of collaboration, or negative lessons to avoid. This type of scan fits both into "activity" and "environment" boxes of the logic model matrix, at the intersection in the Venn diagram below:

Figure 13   **Where an external scan fits into the logic model**

Activity (Team)

External scan

Environment

---
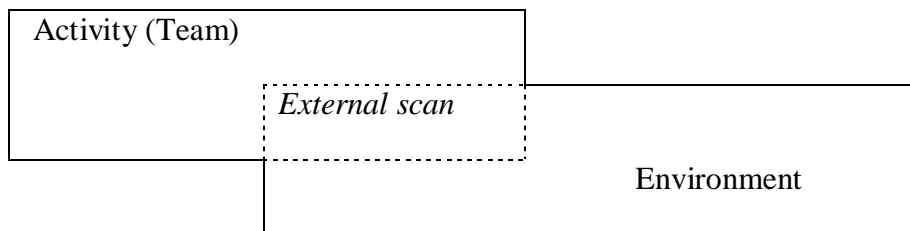
[30] Other techniques that may have enhanced our efforts include Patton's (1990) recommendation to use typologies such as labeling various kinds of data (i.e., P = participant, S = staff) and creating a data index as the first step in content analysis. "Without classification there is chaos. A comprehensive system provides easy access to data for anyone who wants to use it" (pp. 382-3).

External scans had residual gains: Team members reported enjoying networking with professional colleagues at other colleges, creating a global and sharing community of evaluators. Why limit a healthy "culture of assessment" to one campus? I consider the advantages of sharing and collaboration an underlying theme of my dissertation.

Our energy yielded fast results. Without belaboring points or modifying recommendations, we had agreed to act decisively upon reaching consensus. This process would incorporate dissenting votes, too – usually under a "risks" section of a document. The Team's debut evaluation report, on e-portfolios, was completed and submitted to the Dean and distributed to other Student Affairs decision makers in early December 2005. This accomplishment inspired pride that I hope was shared among all Team members. That initial triumph served as a model for future efforts and created positive momentum for next phases and projects. The tone for our next meeting was mildly celebratory.

Preserving momentum, Team members continued searching for useful perspectives on the other current outcome, mentoring. That report was completed the following month, in late January 2006. At the same time, a "staggered" approach allowed us to continue immediately into searching for facts on three other outcomes. This seemed an efficient way to proceed, as we were working actively on something rather than awaiting instructions or feedback. This proactive style boosted Team morale as well as productivity.

Another typical characteristic of our evaluations was the finality of phases. Once we had completed a report, that particular task was no longer in our focus, except on a medium-term horizon when feedback and further instructions would arrive. A logic model with targets and timelines would have helped us keep each "finished" outcome in focus, not only in the back of our minds for future visitation. The Team expected that reaction from decision-makers to

evaluation reports would reflect general management style. So, early on we did not know what to expect. In our case, it turned out that no news was good news – indication that decision makers were satisfied with our output. Our comparison is that other Student Affairs teams received direct messages that more was expected from their efforts. In retrospect, our pace and energy level gave decision makers ample information. We were somewhat concerned that we had set unrealistic precedent for future endeavors. This was a valuable lesson on balancing priorities and managing time within a bureaucracy.

4.1.1.4 Summary of activity for first year of campaign

To summarize methodology, reporting, and findings, the Team's first-year assessment process consisted of the following sequence of steps.

1. Consult source documents, including mission and vision statements and list of Student Baseline Outcomes.
2. Meet as scheduled and discuss issues according to agenda and impromptu.
3. Derive consensus on steps to gather data; conduct investigations.
4. Analyze data; submit findings to Dean.
5. Proceed to next Outcome, determined by consensus.

Those steps culminated in the series of evaluation reports listed in the table below ("Summary of Student Baseline Outcomes evaluations & Recommendations to the Dean of Students for Assessment and Measurement Programs", May 25, 2006, pp. 1-2):

Table 4   **Summary of Student Affairs Assessment Team,**
evaluation reports submitted 2005-06

Outcome report	Release date / status	Outcome #	Comments
Complete			
E-portfolios	Dec. 9, 2005	7	Internal & External scans for ideal models and fit.
Mentoring	Jan. 31, 2006	12	Internal & External scans for ideal models and fit.
Internships Pitt Pathway Pitt Promise	Feb. 9, 2006	8 11 5	Internal scans on current practices in Student Affairs units.
Interviews w/ students	Mar. 22, 2006	2, 3	"Needs assessment" / "needs audit".
Focus Group	May 18, 2006	1, 2, 3	Report on findings; precedent for new focus groups in Fall 06.
Student Affairs performance Outcomes	May 18, 2006	4, 6, 2	Internal scan & status report.
Parents conduit	May 19, 2006	14	Internal & External scans for ideal models and fit.
Pending			
Faculty liaison	active	13	Collaboration w/ Marketing & Communications Team.
Conduct Violations and Alcohol (repeat offenses)	active / summer	9, 10	Baseline data for ongoing comparisons.
Pitt experience	active / summer	1	Analysis & findings of ResLife Quality of Life survey are forthcoming (June 30 target). Key success measures of Student Affairs *Vision*.

Patton (1990) says, "The key issue in selecting and making decisions about the appropriate unit of analysis is to decide what you want to be able to say something about at the end of the study" (p. 167). In the report cited above, the Team shared recommendations for Student Affairs program enhancements, reiterated from individual evaluation report findings. We presumed that observed patterns and themes could guide renewed and ongoing assessment activities the following Fall term. The Team's recommendations may have seemed forthright, despite a disclaimer that "these ideas are not intended to override existing plans or documents, but only to frame possible policy directives" ("Summary of Student Baseline Outcomes

evaluations & Recommendations to the Dean of Students for Assessment and Measurement Programs", May 25, 2006, p. 4). While recommendations for departmental policy identify the Team as an assertive crew, this evidence helps show a balance and range of active evaluation. A less active approach – preferred by some management styles, as I have since observed – might offer only facts and data that decision makers analyze, interpret, and sculpt into modified policy. As noted earlier, recommendations and supporting data can go ignored, too – which Directors have assured is not what happened in this case. The Dean shared impressions that this memorandum represented a valuable culmination of the Team's rookie year. These recommendations represented more than brainstorming, but were instead an encapsulation of year-long investigations, with the following key points.

- Student Affairs customer service model. Focus groups conducted in Spring 2006 revealed that many students experienced frustration in identifying the appropriate source to help navigating the system and to facilitate problem solving. The Team also recommended establishment of a centralized information center, which could take form of a "hotline" or web-based FAQ list.

- Diversity enhancement. A new committee would provide a forum for information and feedback among various groups. Tasks for the Diversity Team would include marketing of programs to the general student population and methods to encourage participation.

- Ongoing focus groups. Based on the results of focus groups, the Team predicted that completion (by Residence Life) of the Quality of Life survey analysis would reveal other issues of interest to student focus groups.

- <u>Alcohol incidences</u>. The Team cited trends of correlation between times of high stress throughout the academic year with increases in alcohol incidences. This recommendation dovetailed with Student Baseline Outcome #9.

- <u>Database of assessment tools</u>. The Team revisited ideas for a feasibility study on a centralized Student Affairs database for assessment tools that would facilitate efficient planning of focus groups, interviews, and training sessions. Such tools might include templates, sample surveys, and draft agendas available through a shared drive or intranet.

The common theme I observe in these recommendations is a reformulation of assumptions, perhaps premature but nonetheless geared towards impact. Upcoming logic model analysis will cover how these recommendations did or did not transition into actual outputs.

Upon completion of the academic year, the Assessment and Measurement Team and all teams went on summer hiatus. Activities among chairs and liaisons included planning communications in mid-August for the resumed campaign.

4.1.1.5 Second year activity

Departmental benchmarking took place in the subsequent academic calendar phase. Ownership of assessment transitioned to units, as the Dean expressed confidence that the envisioned "culture of assessment" had begun to take root. Among the annual summer kick-off workshops were sessions instructing Directors how to manage unit-level evaluations. (I was not privy to these sessions, nor granted access to minutes and other supporting documents.)

We awaited instructions to re-convene. Assessment Team Meeting Minutes from October 24, 2006, say that the Team commenced our first "official" meeting of the season with a discussion on appropriate framings for upcoming assessment activities. Consensus was that process should follow up on last year's outcomes with fresh status appraisals or audits in all

Student Affairs units. Re-measuring baselines was perceived as a useful activity to renew our campaign. In my opinion, this type of effort could involve clarifications of certain outcomes' wording, by the Dean's instructions or concurrence. For example, reminding students of the Pitt Promise is better described as process than outcome. I think this observation reveals that an effective leader's understanding of evaluation nuances need not be perfect, only authentic and well-intentioned.

From archived documents, the chairs created new templates for Student Affairs performance outcomes measured the prior spring. This task was an effort to create more formal and durable instruments than simple e-mailed three-question inquiries to gather information. Another source model was our year-end assessment report. For units that did not submit formal reports, results and "lessons learned" were gathered through other documents and correspondence. From these models and combinations of techniques, we aimed to recommend "best" practices for certain situations.

As "completists" (arguably a positive attribute for an evaluator), we planned to revitalize unfinished evaluations of pending outcomes on alcohol and conduct violations. We determined a favorable course of action to collaborate with Judicial Services (USJS) and the Fraternity & Sorority Council. Preliminary observations indicated that alcohol violations were "down a tad" from prior year, while illegal downloading had risen. We sought other data sources, primarily gathering students' perceptions of PEAR (Personal Education, Assistance, & Referral) classes on alcohol use. Having conducted a class on alcohol awareness, a consultant in Student Health was considered a valuable contact. (Brief analysis of the alcohol policy outcome awaits ahead in Appendix B.10.)

As we compiled research on methodologies active on campus, we continued to analyze and compare modalities and culture at peer and aspiration institutes. This information, along with analysis of our own prior evaluations, enabled us to envision a "checklist" of components in thorough assessment. Also, it prevented us from wasting effort "reinventing the wheel". While we recognized Michigan as a prime aspiration institute, a team liaison was entrusted to maintain and distribute a complete and updated list of benchmark universities.

Once we had updated a list of benchmark institutes, we realized we could draw comparisons several ways. Most compelling were 1) by overall Student Affairs performance per institute, and 2) per unit across institutes (i.e., seeing how Student Health conducts evaluation at Rutgers and Illinois). Choice of approach would trigger how the Team would structure an external scan and divide the labor.

Our final discussion at the season's introductory meeting involved a review of the Student Affairs FY07 Long-Term Goals document. Some of these goals were new, and some re-formulated from last season. Among those specified goals, we saw our primary directive to continue developing a "culture of assessment" with resolution by "benchmarking data… to provide a baseline for various programs in Student Affairs" (Assessment Team Meeting Minutes, October 24, 2006, p. 2). These directives demonstrated that mission had not changed, an essential framing for logic modeling[31]. This was a promising sign that management continued to trust the Team's judgment and output.

The Assessment Team recognized different sets of outcomes driving the activities and attention of other formal teams and task forces. Recognizing that collaboration with other entities would share knowledge and limit duplication of effort, Team chairs occasionally attended other

---

[31] A follow-up study might explore how a logic model must adapt if mission changes.

teams' meetings. As "amicus curiae" I attended meetings with the Leadership Committee. To promote reciprocity and effective communication between Student Affairs units, we invited that committee's chair to an upcoming meeting. Some delegates joined more than one team and were able to serve as messengers.

One ironic result of the Assessment & Measurement Team's remarkable productivity in our initial campaign was finding ourselves in a holding pattern as Directors, according to Team suspicions, may have been catching up with the vast amount of information we had delivered. A primary lesson was to pace ourselves. While the Assessment Team did not receive further marching orders, we continued to meet monthly to stay on track and maintain positive energy. Assessment Team Meeting Minutes from November 21, 2006 address "a recognition to wait for official notification of any reformulated statements and interpretations of outcomes, and instructions how to proceed" (p. 1). A logic model with target dates and defined phases could perhaps have enabled manageable, efficient pacing.

A pending outcome from the previous year's campaign dominated the remainder of discussion at the November 2006 meeting: "The Division of Students Affairs will provide Faculty with additional tools to better understand and assist students." Brainstorming resulted in a list of various tools and techniques to accomplish this goal:

- directory of contacts
- quick reference card (laminated)
- web page, including FAQs
- orientations
- handbooks
- written communications

This list would populate an "output" box in a logic model.

The Team sage offered warnings that the goal was not to create or eliminate programs but to change indifference and enhance utility. Popularity of such utilities among faculty would

81

depend on their specific needs, not necessarily Student Affairs needs or perceptions. Current practices included a new newsletter produced by the Student Affairs Communications & Marketing Team. One of our tasks would be to measure its use.

Conversations resulted in two-pronged activity. First, an internal scan would reveal what was happening concurrently in general unit-level assessment. As Directors were already conducting per-unit benchmarking, Team members would request that his or her Director provide a description of actions within each department to equip faculty with necessary tools and knowledge. Next, an external scan of aspiration institutes could reveal favored and current practices. The Team considered conducting a survey via websites and personal communication by phone or e-mail with Student Affairs contacts at selected schools. Despite sentiment that surveys of other peer institutes could add value, we concurred that strict focus on aspiration institutes could likely reveal favored practices, make efficient use of our time, and conform to Pitt strategic directives. Team members chose assignments and we established a target deadline for the following month – in time to offer an update at the upcoming Quarterly meeting.

In the second year, our Team produced only one new evaluation report, on Faculty tools, submitted on February 6, 2007. This may illustrate the dangers, if they may be called that, of embarking upon a new campaign with such energy that team members find themselves comparatively idle the following campaign. Rather than looking for work or creating work, it made sense for us to sit tight while management caught up with our prior activity. At the same time, it was important to keep on schedule with monthly meetings full of updates and new material, some based upon contemporary research, conference findings, even administrative gossip about the assessment process. That way we were always poised to jump back into action. I reiterate that perhaps a logic model could have enabled us to pace ourselves.

4.1.1.6 Concurrent unit-level evaluations

Team activities comprised only a portion – arguably the most visible one – of the department's overall assessment efforts.   Within Residence Life and other departments, independent evaluations took place on a regular basis. While these projects were part of an overall picture of university program and academic assessment, **this study focuses on the Team activity centered around Student Baseline Outcomes**. The diagram below illustrates how unit and team level evaluations fit together, using Residence Life as an example. *The green box represents the scope of this study*; every other element in this diagram represents a related but separate topic serving Student Affairs vision of a "culture of assessment".

Figure 14   Residence Life evaluations as subset of Student Affairs assessment process
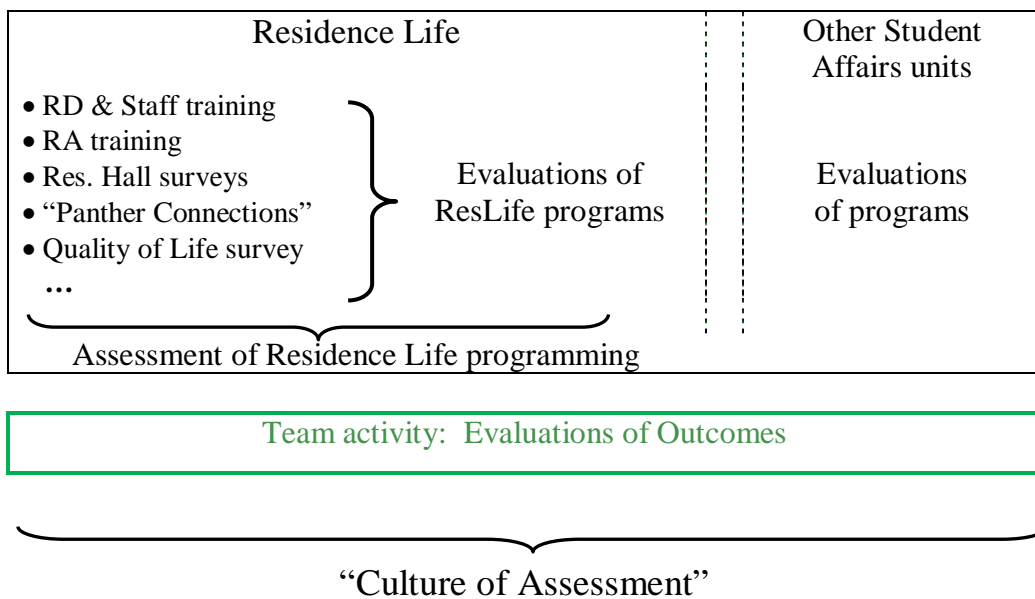
Informal interviews[32] with unit directors revealed that some kind of measurement and benchmarking took place in every unit concerned with customer service and improvement. Unit-

---

[32] These interviews amounted to me asking each director at the December 2006 quarterly meeting, "What kinds of assessment and benchmarking are you doing in Career Services…? … in Student Health…?" , etc. and receiving an

level evaluation projects often demonstrated a quantitative approach bolstered by qualitative follow-up. For example, students completed Quality of Life surveys in the spring, shortly before classes ended, and returnees were invited to focus groups or interviewed about Student Affairs topics that deserved greater attention according to survey results and analysis. Some of these projects were confined to the unit, while others such as the Quality of Life survey shared accountability and were reviewed among higher management before delivery. Collaboration and accountability depended on each project's scope and overlap between units, but all Student Affairs staff were expected to be aware of events and policies throughout the department[33].

4.1.2 **Post-service reflections**

The following summer I changed jobs and lost direct access to this wealth of data. My standing went from emic to etic. This was an interesting change in perspective, similar to that of evaluators contracted for a project without sharing a formal stake. Any data I might seek to include in this analysis would come from hearsay and interviews rather than from immersion in and observation of the process. If analysis of subsequent phases were in scope for this study, my approach might feel urgent as my access to fresh data had become tenuous. This transient condition has given an important lesson on the contrast between insider evaluators, who retain an ongoing stake and access, and outsider evaluators who may be confined to a time frame and availability of quality data. I feel that these contrasting viewpoints represent two separate dissertations – and perhaps successive chapters of an eventual book.

---

affirmative reply that either the director or a delegate was doing that. I unfortunately have no transcripts of these off-the-record conversations, only personal hearsay.

[33] For example, the Director of Residence Life was expected to be competent and informed enough to substitute for the Director of Student Health, and vice versa, if circumstances demanded.

Within Student Affairs, philosophical and operational adjustment was that the Assessment and Measurement Team went from being an advisory panel on assessment process to also being an advisory council for other teams. Underlying assumptions were that 1) mechanics of assessment had moved to the unit level, with unit / Director accountability; 2) Student Affairs management and planning had elected to follow a dual-pronged approach whereby a voluntary advisory body would share unit-wide assessment accountability with formal organizational structures; and 3) all "team" functions (i.e., marketing, communications, staff recognition…) would integrate with assessment.

4.2   INTEGRATING CASE STUDY AND LOGIC MODEL

The previous sub-chapter told what happened – the facts of the case. The signature elements of Pitt's campaign to create a culture of assessment in Student Affairs were determination of Student Baseline Outcomes and a designated Team's leadership in investigations.

This sub-chapter's task is to meta-evaluate evaluations of those outcomes. In response to my first research question, demonstrating continuity and identifying gaps in linkage or implementation between assumptions and outcomes by retro-fitting the case study into a logic model reveals the wisdom and success of the various evaluative approaches taken by decision makers and support teams. My premise is that satisfying a logic model metric constitutes thorough evaluation of that outcome.

Heeding advice on scope (sub-chapter 3.1.1) and intuition on narrative economy[34], I limit meta-evaluation of the Team's activity to three selected Student Baseline Outcomes, representing the Team's first effort and, in my opinion, its best and weakest efforts. (Summaries of the remaining outcomes, which contribute to overall ratings and findings, appear in Appendix B.)

The figure below, a restatement of Figure 3, illustrates mapping of evaluative elements, factors, and processes into the logic model.

Figure 15   **Logic model depiction of assessment cycle**
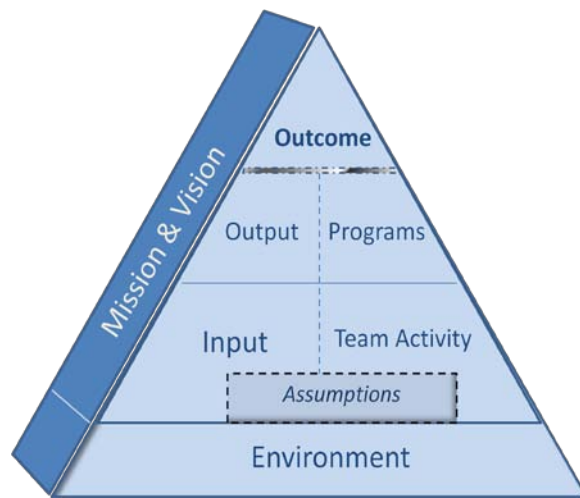
The model features mission and vision as the "spine". Actualizing a "culture of assessment" is the unit's proclaimed mission. Along with vision "to provide University of Pittsburgh students with the best collegiate experience in the country," this mission drives every intended outcome. Each analysis sub-chapter begins with outcomes retro-stated as criteria. For component elements, it is important to heed Patton's (1990) and Weiss' (1997) advice that plugging variables into columns is not formulaic and rote, but inductive – that is, reflecting causal relationships.

---

[34] In an exhaustive analysis, I would investigate evaluations of all 14 outcomes chronologically, in the same sequence that the Team proceeded, as to preserve logical progressions between outcomes as perceived by the Team. Such bulk better suits a textbook manuscript than dissertation.

Next, I rate quality and thoroughness of Team performance on individual outcomes by fitting "scores" into each logic model slot, according to the following metric of symbols:

Table 5   Key to logic model "scores" in meta-evaluation

Symbol	Meaning
✓	Element has been analyzed adequately.
✓+	Team did an exemplary job in this task.
( )	Element is unresolved or conditional.
-	Element or task was not resolved.

In response to my second research question, I appraise how the Team did and what it might have done with knowledge of logic models. Did Team actions add up to comprehensive assessment? After each logic model gets filled with status marks (check, plus, minus, parentheses) denoting qualitative judgments, I translate and tabulate ratings into total scores. That complete tally is compiled into Table 7, forthcoming on p. 120. That table provides basis for meta-evaluation of the whole campaign.

Instead of analyzing all 14 outcomes, I have selected three representative outcomes. First is the Team's debut evaluation, on e-portfolio programs.

4.2.1   e-portfolios (Student Baseline Outcome #7)

The Assessment and Measurement Team's first completed evaluation report, on electronic portfolios, was released to the Dean on December 14, 2005. In pursuit of resolutions to this study's two research questions, this sub-chapter uses the study's featured logic model to analyze how well the Team conducted its evaluation of e-portfolio programs, and speculates how the Team might have performed had they used a logic model. Analysis follows the layer-by-layer

approach suggested by the logic model's shape, and commences with the intended outcome restated as criterion.

Outcome (Intended): "Student electronic portfolios will be investigated to determine feasibility."

Environment. Through an external scan, evaluators can locate models and precedent that may fit with the host institution's profile (House & Howe, 1999; Upcraft & Schuh, 1996). I feel that the Team's assessment of environmental factors was comprehensive and accurate. The Team did an admirable job with the external scan – an exercise that satisfies the "environmental" scope of a logic model.

The Team found that two aspiration institutes reported success with outsourced e-portfolio platforms. Advantages to outsourced agreements included expertise, efficiency, access to wide databases and comparisons, and quality. Also, outsourcing would allow university administrators to focus on core competencies without taking their own time to create connectivity networks or to oversee dedicated staff entrusted with those tasks. Disadvantages included cost. To avoid such costs, an aspiration institute created e-portfolios in-house. Longer-term cost-benefit analysis could determine if outsourced costs would exceed salary and benefits for additional in-house staff dedicated to the same tasks and maintenance. In-house costs would be minimized or subsidized when the university could enlist active technical staff to take on an additional project on e-portfolios. Another university secured a federal grant to cover costs for e-portfolio networks.

Findings included a roster of institutions currently using electronic portfolios ("Evaluation of electronic portfolios", December 14, 2005, p. 4):

- Illinois State University, Center for Teaching, Learning, & Technology;
- Ohio University, Career Search;

- Kalamazoo College, whose Portfolio was funded by a grant from the Fund for the Improvement of Post-Secondary Education (FIPSE);
- University of West Florida, College of Education;
- University of North Carolina, who outsourced "Optimal Resume" with reportedly "excellent participation" from both students and employers;
- University of Virginia, who outsourced "Interfolio.com", available at a cost to students/alumni, and mostly used by alumni (~3,360). University representatives said, "It's the best decision that they have made."
- Wisconsin, with a system developed in-house.

This list represented aspirant institutes from Pitt's official list, and others who would fit a profile analogous in some respects to Pitt. While the list included no "peer" institutes from Pitt's official list, the Team found concurrence among four peer institutes and two aspiration institutes who were <u>not</u> currently using e-portfolios: Penn State, Maryland, Michigan, Buffalo St., Michigan St., and Rutgers. The next step (to my knowledge, not yet pursued) would be to find out why they were not used there. Any unfavorable testimony might give planners pause for continuing the project and warnings about how to proceed most efficiently. Some Team members speculated that other schools' abstinence from formal e-portfolio programs supported counter-arguments that customized, personal help – perhaps from Career Services – represented a sensible and efficient use of funds, time, people, and space more than a formal e-portfolio system.

Environmental pros and cons under discussion included infrastructure, competing priorities, and support for and compatibility with a new human resources technology package scheduled for imminent implementation. Other environmental factors included comparative and opportunity costs; institutional culture and welcoming outsourcing versus in-house or organic experts and trainers; and efficiency.

In its report, the team recommended that Pitt could consider agreements with outsourced consultants or in-house dedicated staff (p. 2):

- Corporate partnerships or sponsorships.
- Collaboration with the School of Information Systems or the Katz School of Business. Such arrangements might include "insourced" internships.
- Collaboration with other institutes. Do outsourced consultants consider packaged agreements or references (for discount)?

Each of these environmental solutions aimed to deflect or share costs while boosting efficiency with pooled knowledge.

Inputs: *Student attributes*. The Team recognized in its report an essential preliminary step, before Pitt would sink time and money into research and agreements: a thorough needs assessment, to determine whether undergraduate and graduate students and alumni would use this service. What are students' incentives? Could participation be encouraged by granting academic credit for participation and use? The Team advised policy designers to determine a representative sample, mainly juniors and seniors thinking about future employment. Casual interviews and hearsay from RA's indicated that many freshmen impulsively replied that they would definitely use an e-portfolio platform or any technological service. A survey to random residence hall floors may have resulted in an accurate measure of student demand for e-portfolios. In its report, one of the Team's foremost recommendations was for a thorough initial student needs audit for such a product or system.

Assumptions. Planners recognized that this generation of college students is facile in technology, and would possibly take a shine to an electronic utility. However, no formal internal audit was conducted other than a word-of-mouth gathering campaign by RA's. I suspect the Team made a rookie mistake in deferring to assumptions over thorough gathering of concrete details. The Team speculated correctly that periodic seminars might encourage less adept students to overcome fear or hesitation towards technological solutions. As happens with many other programs, whether instructional or leisurely in nature, the Team saw value in enticing

participants with food or give-aways, such as T-shirts. Another audience considered difficult to reach were "entitled millennials accustomed to praise for minimal effort and independent survival skills" (December 6, 2005 meeting minutes, p. 2). We found substantiation in literature for these attitudes observed and reported to us by RA's and RD's from informal internal scans.

Assessment Team activity. Perhaps because we did not perceive our work as an academic mission, the Team did not conduct extensive or intensive scholarly research on this or many other topics. Instead, brainstorming substantiated points of concurrence with peer comparisons online or testimony from decision makers. While the Team did not presume that management was better versed in the literature – possibly true – we accepted their opinion as decisive. In retrospect, a scholarly portion of evaluative research could supplement the "quick and dirty" approach that was recommended.

The Team borrowed from Barrett (2001) an online definition for an electronic portfolio: "a collection of documents and artifacts using on-line and interactive technologies, allowing the developer to organize information in many media formats – including audio, video, graphics, and text – and to create linkages to other portfolios, sources, and references". While our report cited artistic, financial, and educational portfolios as example types, Barrett and others cite primary purposes as "Learning (Formative) Portfolios, which usually occurs on an ongoing basis supporting professional development; Assessment (Summative) Portfolios, which usually occurs within the context of a formal evaluation process; and Employment (Marketing) Portfolios, which are used for seeking employment" (Barrett, 2001; Hartnell-Young & Morriss, 1999; Wolf, 1999). The Team's focus was upon portfolios best described as serving "learning" and "employment" purposes, which the Team assumed to have the greatest utility and application for Pitt students, alumni, and partners.

Perceived technical and utility benefits included the following:

- minimal storage space
- ease to create back-up files
- portability
- long shelf life
- learner-centered character
- increase in technological skills
- through hypertext links, ease to ensure that certain standards are met
- accessibility (especially web portfolios)

While possibly accurate and thorough, this list had one flaw in uncertainty whether its source was presumption or direct inquiry of students.

More compelling advantages included environmental factors tying into Pitt infrastructure and cost-benefit analysis, with learning and technological skills as student input factors. The Team concluded from an internal scan that Pitt had the infrastructure to implement e-portfolio platforms. Moreover, Student Affairs had the budget to set up systems, allocate storage, and set up contracts with professional services and consultants for training. Student Affairs had staff experts willing and eager to serve as mentors and trainers. Not coincidentally, this factor ties in with the mentoring outcome evaluated next by the Team, indicating we understood possible efficiency advantages of bundling outcomes.

However, preliminary (summative) findings from Career Services experiences put into doubt the value of e-portfolios to employers, who may lack time and energy to analyze such an expansive document. Still, the Team considered that e-portfolios may be useful for graduate job-seekers in honing their interviewing, organizational, salesmanship and presentation skills, along with enhancing their confidence.

To balance and triangulate points, the Team consulted colleagues at Virginia and other aspirant institutes who had effectively implemented e-portfolio systems for undergrads. Testimony amounted to hearsay from administrators with perhaps a vested interest, but they

claimed that e-portfolios were well received by the student body and resulted in higher numbers of job offers and graduate school contacts. (A separate study / literature review could verify and quantify these claims, accepted on good faith by Pitt Student Affairs.) In other words, an external scan revealed that colleagues at respected universities shared Pitt Student Affairs planners' belief in this outcome and that it could be trend-setting for post-graduate planning and employment, likely resulting in greater alumni satisfaction and loyalty – often a bottom line for any directed outcome.

In its December 2005 report, the Team advised decision-makers, in considering implementation of one or more e-portfolio programs, to heed the following issues:

1. Examining the "portability" of e-portfolios – users' access after graduation. Often an alumnus would pay a third-party vendor for that service. As precedent, "Interfolio" at Career Services cost an estimated $35 to $45 per year, deemed affordable even for students on a tight budget.

2. Another possibility is to post a "how-to" template on the Career Services website, or the utility page in the my.pitt.edu portal.

3. Are e-portfolios best constructed through an outsourced agreement and software packages, by creation of in-house expertise and technical support, or organic and independent creation by students (now the prevalent "system")? What arrangement would work or already does work best in comparable contexts and most economically at Pitt?

The Team identified an alternative approach that Student Affairs might not need to centralize this initiative. It may be more efficient that each Academic Department or School could autonomously create on their websites an e-portfolio "bank" or customizable template for their students – instead of making such a template available only from Career Services. This

arrangement may better accommodate each academic / professional field's own format of portfolio. Also, unit "ownership" could reduce possible burdens on Career Services in terms of budget, personnel (for instructing students on the construction, use, and maintenance of e-portfolios), and administration (i.e., website management, troubleshooting).

The Team recommended continuing to investigate useful models for other pilot programs with contrasting structures, perhaps in the Honors College or within another new living learning community (LLC). This would set up sample groups with e-portfolios and allow contained, measurable tracking without launching into an expensive campaign to equip the whole student body with this utility.

Outputs: *Programs*. The Team located precedent in the "Student Development transcript", established in 1979 and scrapped around 1985. This service received limited usage, primarily from top-notch scholars who were equipped for competitive employment – with or without a fancy presentation of their resumes or CV's. Some Team members thought this program gave further evidence that an effective e-portfolio program "should empower users, rather than do [the job] for them ("Assessment Team Meeting Minutes", December 6, 2005, p.2).

In recent precedent, the School of Education's quarterly newsletter reported receiving a grant to develop an e-portfolio system to support preparation of teachers and school leaders. A Team member located evidence of a prior evaluation of e-portfolios for the Athletic Department, with a pilot program of ten scholar-athletes. The Team determined that Pitt had precedent in the Katz Business School, School of Education, and Information Services on effective and popular e-portfolio systems, although percentages and statistics were not precise. The Team recommended a follow-up to seek collaboration on that project and share information.

94

Precedent, assumptions, and reporting yielded a quick result. Looking forward from the Team's December 2005 recommendations, a pilot program for e-portfolios was launched in a residence hall LLC in the fall of 2006. Absent formal feedback from the Dean of Students (who understandably withheld certain details of executive privilege), the Team later shared a sense of accomplishment that our report directly influenced the nature and speed of that project's implementation. If that is true, then here evaluation has served intended purposes.

Outcome (Actual). The team saw some evidence that the previous year's report informed new policy, as Student Affairs launched pilot e-portfolios in Career Services, in the Research LLC, and in partnership with the School of Education. This action represented purposeful collaboration between management and evaluators.

Impact. Future evaluative steps embedded in the logic model could determine programming causality: whether students benefit from e-portfolio programs in terms of career opportunities, networking, technological facility, and general well-being. The next phases of mapping inputs and environmental factors to the intended outcome can diagnose successes and shortcomings of the e-portfolio program. One way to validate this supposition is to interview management on causality between evaluation findings and directed implementations. Causality might range from being a contributing factor, to a decisive factor in supporting one policy directive over another – including an option not to modify policy or to table that decision. Surely participation is an essential measure: calculating how many students actually use this program, either by an electronic survey – possibly one or two questions in the annual Quality of Life survey – or by polling RA's.

<u>Meta-evaluation</u>.

Figure 16   **Logic model meta-evaluation of e-portfolio outcome**
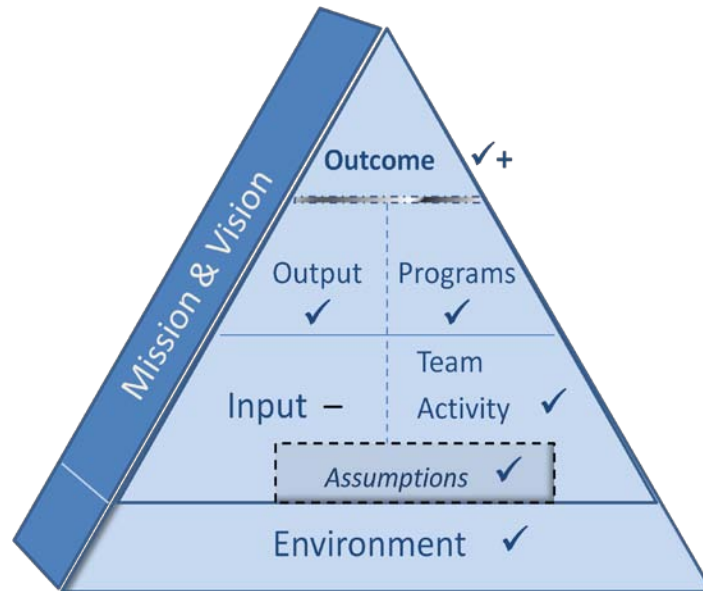
From the onset, this directive was partially driven by necessity. The Team had to get started, and picked an outcome that seemed straightforward, rather than perhaps another outcome that was most urgent. The Team was getting its feet wet. The Dean had bravely delegated the Team authority to choose its own path, and so we ran with this outcome.

The Team did a thorough job investigating this outcome and making productive recommendations that have resulted in new policy. All of the blocks in the logic model contain information. However, the Team could have done a better needs audit to determine student desire for and perceived benefit from this outcome. Some of the process serving this outcome was left to assumption and consensus rather than defined and measurable demand and gain. Had the Team actually used this model, I believe that the process would have been more organized and less improvised.

I award the actual outcome a check-plus, because not only did Student Affairs achieve its intent in investigating feasibility of e-portfolios, but also implemented new e-portfolio systems. The next steps – depicted as flow of knowledge back down-slope from the tip of the pyramid to assumptions and activities – must be to assess the endurance of these programs by conducting interviews and focus groups with students to determine their reactions to and use of e-portfolios. The logic model demonstrates how findings and outcomes cycle back to reformulated assumptions and policy.

Figure 17   Logic model depiction of "impact" as knowledge flows from outcome to assumptions
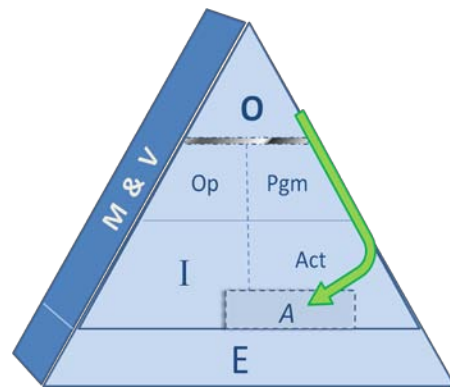
In this case, meta-evaluation could promote a possible re-invention or interpretation of the outcome itself. One criticism of the phrasing for this outcome is emphasis on process and research: "Student electronic portfolios will be investigated to determine feasibility." This is a declaration of summative evaluation, but a bolder statement of outcome might emphasize the importance and gain of e-portfolios – say, "E-portfolio programs will be implemented in key academic domains by the end of the next academic year." Compelling sub-text could assert that an e-portfolio would allow potential employers and graduate school recruiters an easily accessed and concise encapsulation of every Pitt student's academic and extracurricular career.

While feasibility is a critical consideration, I think the key factor to determine this outcome's success is student use. Without that evidence, impetus might have been a perception

that e-portfolios were "the way to go", rather than a necessary outcome on its own merits and based on a need expressed by Pitt students. The Team unearthed no solid evidence from Career Services or other inside sources that students had cried, "We want e-portfolios! Bring them now!" This outcome was at least partially motivated by politics over logical process flow. As the December 6, 2005 meeting minutes said, "A sensible initial step might be a more thorough needs assessment" (p. 2).

➔ Analysis using this logic model has revealed gaps that I and other analysts did not perceive during our prior evaluations. Accordingly, the model points towards future steps of assessing utility of e-portfolios: How many students use and enjoy e-portfolios? How have they benefitted? Are costs justified? Next steps should attempt to answer those questions about impact, which informs input variables for future evaluative phases.

4.2.2 Global learning (Student Baseline Outcome #4)

This outcome represents a meager evaluation by the Team – something I did not realize until using a logic model to analyze accomplishments and shortcomings in evaluation processes.

Outcome (Intended): "Students will have many opportunities to gain diverse and global learning experiences."

Environment. Had we known about logic models, perhaps we would have started with an external scan. Instead, the Team launched into investigation of other factors without returning to environmental framings. Tapping into available knowledge on global programming would have shaped context and determined types of opportunities that might have appealed to Pitt students. We did not conduct a complete external scan for current wisdom on various countries, including safety, currency exchange, language, customs, and other information that might identify most

likely and eager partnerships. Sources of knowledge could include blogs, travel logs, and international programming directors at other campuses. The Team also did not explore partnerships with other colleges, either regional (i.e., Duquesne and CMU) or peer (Rutgers, Syracuse) for information on exchange programs that might result in economies of scale and shared resources. The logic model – with a big, empty "E" slot – could have alerted us to a considerable missing piece.

Inputs: The phrasing of this outcome might provoke immediate and deep investigation of student input attributes, including demographics, international backgrounds, and interests in study and travel abroad or special service projects (i.e., Peace Corps volunteering). As we started directly with the "source" at Career Service to evaluate Pitt Pathway[35], why did the Team not tap into International Services, which catered to students of diverse backgrounds? One explanation for neglect of input variables was our lack of a framework, such as a logic model.

Assumptions.   The phrasing of this outcome presumed 1) diversity and global consciousness each a good in itself, and 2) that interested students would seek out opportunities. The outcome does not step beyond a simple supply / demand statement, which is less forcible and paternalistic than saying, for example, "[All] students will take advantage of opportunities to study abroad and increase global consciousness." The task force who formulated outcomes may have been deliberate to avoid such pronouncement that could echo academic distributional requirements (i.e., being "forced" to take a foreign language class).

Assessment Team activity.   The Team's evaluation of this outcome was in tandem with two other Student Affairs performance related outcomes, with findings compiled[36] into a report

---

[35] Analysis of Pitt Pathway awaits next sub-chapter.
[36] I do not believe that combining investigations of multiple outcomes violated a "lump" effect that Weiss (1997, p. 243) warns against.

submitted on May 18, 2006. Assessment Team members inventoried their respective units on what programs were in place to meet this outcome for "diverse and global learning experiences".

Outputs: *Programs*. The recommended audit identified the following active programs, geared toward diversity and global opportunities.

• Student Life supported 30 cultural and nationality organizations that provided an array of programming. Such affiliations included student organizations for Black Students (BAS), Asian Students (ASA) and Women (CWO). Other student organizations embraced "diverse" art and performance. Programs included art shows, foreign films, speakers and participation in International Week. Student Life promotional materials claimed that diversity had been a standard topic of student leadership training programs. After diversity related experiences and outcomes had been identified, focus groups would follow – also tying in neatly with Outcome #3 on "listening".

• Residence Life identified the following programs promoting and enhancing global and diverse learning opportunities:

a) The International student living and learning communities (ISLLC) hosted bi-weekly seminars, with focus on globalization.

b) A variety of campus-wide and in-hall programs on diversity and multi-culturalism allowed RA's "considerable autonomy to design and customize programs" (Evaluation of Student Affairs performance outcomes", May 18, 2006, P. 2). RA training class devoted a full week to diversity.

c) Beyond programming, the Student Conduct system encouraged students to develop into "citizens of the world" (p. 2). This system was designed particularly to influence students who had exhibited anti-social or offensive behavior, according to notions of community

standards and personal identities. This program integrated with the "conduct code violations" outcome.

d)   Residence Life served student needs related to diversity and global awareness through the Quality of Life survey, informational Floor meetings, flyers, mass and targeted emails, *Pitt News* advertisements, and program publicity. Interaction included one-on-one meetings with RA's, who were "trained to listen well and to report back to directors" (p. 2). In general, Residence Life operated on the fundamental principle of serving student needs, including referrals to other Student Affairs units better equipped to answer a question or address an issue on diversity and global awareness.

e)   Residence Life partnered with Pitt Arts (a sub-unit of Student Life) to host a weekly Artful Wednesdays noontime performance series in the lower lobby of one of the freshman residential towers. Performances included a reggae band with members from Jamaica, a Polynesian dance troupe, an American-British baroque group, Latin American flamenco guitarists, hip hop, and countless other memorable sessions. Performances were generally well attended[37], and not only because of the free sandwiches and cookies.

•   Student Health supported more than 100 programs throughout the year targeted to wide and diverse audiences. Programs affiliated with student organizations involved the Greek community, residence halls, athletics, and Student Government Board. Topics included diversity, in addition to the expected sessions on meningitis, first aid, CPR, drug and alcohol awareness, and stress management. Patient education materials were available in several foreign languages. Other specific programs included "eat and meet" sessions intended to encourage students of various ethnicities to mingle and share impressions.

---

[37] In fact, I regularly attended both for interest and support.

• Judicial Services (USJS) did not have programming geared towards diverse and global learning experiences – or, for that matter, any formal programming. As with several outcomes, Judicial Services was an exception to active programming, because of the disciplinary and sometimes confrontational nature of its purpose. USJS illustrated how different areas might be held to different standards in an otherwise holistic and unified assessment campaign.

• Disability Services (DRS) collaborated with the Office of Study Abroad to support program participation by students with non-apparent disabilities (i.e., reading disorders, psychiatric disorders, chronic health conditions, and Attention Deficit / Hyperactivity Disorder). Disability Services referred students to the Office of Study Abroad to explore opportunities and make necessary accommodations to prepare for potential obstacles while abroad. Disability Services met annually with Study Abroad staff to update one another on processes, opportunities, and disability related concerns. Disability Services and Study Abroad secured funding through inclusion in the Sissy Lieberman Scholarship, a fund for underrepresented students to study abroad. Students who received a financial award were required to keep travel journals and discuss their experiences at an annual luncheon. In the prior two years, more than fifteen students with documented disabilities had participated in study abroad experiences. (I would award this program a gold star for its breadth, practicality, and purpose.)

• International Services (OIS) specialized in this outcome by providing multiple opportunities for "diverse and global learning experiences" to international and domestic students, particularly celebratory programs bringing together constituents:

a) OIS Afterhours was a monthly social event designed to help international students and scholars discover local social and cultural resources. This program provided an opportunity

for internationals to communicate with, visit, and interact with American social, historical and social institutions.

b)   International Coffeehouse was an organized party showcasing aspects of home cultures to Pitt affiliates. Cultural features might include cuisine and drink, costumes and traditional garb, pastimes, music, movies and entertainment. Interaction between participants and hosts was encouraged through activities such as board games, cooking demonstrations, art and folk crafts, displays, and demonstrations. Entertainment included movies, slideshows, fashion / costume shows, music and dance performances.

c)   The award-winning International Demonstration on Cooking allowed students and scholars to engage in conversation over the universal theme of food. By drawing participants from over 20 countries, the program promoted regional and international culinary pleasures as common and unifying "languages", while providing a safe environment in which to practice English conversation and presentation skills. The small program size (limited to 14) facilitated one-on-one interaction between individuals of diverse cultural backgrounds ("Evaluation of Student Affairs Performance outcomes", May 18, 2006, p. 3).

d)   International Week at Pitt promoted more than 30 activities and events during a week in September designed to expand awareness of and interest in global learning opportunities. Program designers said, "The program celebrates the intercultural diversity of campus life and promoted the University of Pittsburgh as a center for international education" ("Evaluation of Student Affairs Performance outcomes", May 18, 2006, p. 4). Events were sponsored jointly by OIS, the University Center for International Studies (UCIS), the Area Studies Offices, academic departments, and student organizations.

I add that the existence of a functional Office of International Services – let alone a vibrant one – *prima facie* satisfies the "program" portion of the logic model.

• All opportunities that came into Career Services (internships, part-time jobs, full-time jobs) were listed in the PantherTRACS database, accessible to all Pitt students and alumni. From this data, the Emerging Leaders program strove to recruit a diverse body of participants that reflected the campus population. Career Services claimed specific focus on valuing others and exploring diversity within this program.

The Associate Director in Career Services worked collaboratively with OIS and UCIS to offer programs relevant to international and American students who were interested in working abroad. The AD worked with several schools and departments across campus for subscription to "Going Global", an internet resource that informed students about opportunities overseas.

• The Counseling Center promoted global and diverse opportunities with multicultural seminars and an international open house in the Fall season. The Center had conducted a needs audit mainly through focus groups, surveys, interactive therapy process, communications liaisons with Residence Life and academic departments, and prompt response to inquiries. Support groups and workshops were developed throughout the term, and program designers reported that themes were related to diversity and global awareness ("Evaluation of Student Affairs Performance outcomes", May 18, 2006, p. 5).

Outcome (Actual). *Results.* The Team's audit revealed that Pitt Student Affairs units actively engaged in many programs intended to enhance students' global and diversity-centered experiences on campus. The Team's report on May 18, 2006 compiled baselines upon which each unit was encouraged to seek improvement independently and collaboratively. This collaborative process would likely center around communication, including a sharing of effective

104

practices and policy that had or had not resulted in intended goals. For example, elimination of unpopular programs on diversity (such as the "canned" lectures in RA summer training) might allow units to redirect funds into tested and possibly even experimental ventures.

The Team recommended that next steps should involve monitoring progress from newly established baselines. Techniques should include a combination of focus groups, interviews, surveys, and directors' instructions. Student Affairs planners could add content to the "environment" block of a logic model or other assessment framework by seeking closer partnerships with the Pitt Study Abroad Office and with other colleges seeking similar development. A great opportunity to integrate certain outcomes would be to make global and diverse opportunities one "station" or checkpoint in its vision of Pitt Pathway[38].

Unfortunate irony was concurrent dissolution of Pitt's Semester at Sea program. Good news was that the Study Abroad Office has revitalized programs to offer students "first-hand experience that will serve them in an increasingly global environment" (Barlow, 2008b, p. 1).

---

[38] Analysis forthcoming, so this claim will make more sense then.

<u>Meta-evaluation</u>.

Figure 18   Logic model meta-evaluation of global and diverse opportunities outcome
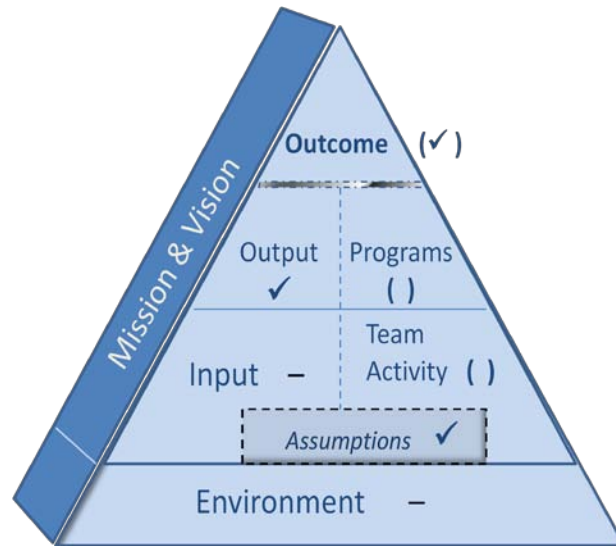
The Team's research established that this was an outcome where progress was already happening. Students were actively engaged in dialog on campus and in programs abroad that satisfied this outcome's parameters. So, the Team's significant accomplishment on this outcome was a thorough inventory of active programming – helpful knowledge for planners and project leaders perhaps considering summative options. So the Team earns a check mark in the "output" slot. While no new programs had resulted during the scope of this study, I hope that descriptions under the "activity / programs" heading have offered the reader a closer glimpse of Student Affairs programs' operation and tone – reasons why students might want to attend. Existing programs were good ones.

The Team neglected an environmental scan that would indicate what peers and aspiration institutes were doing to accomplish similar goals. A possible boon of that research would be collaboration between schools. How about an ongoing joint program with Syracuse or Maryland to send student teams abroad for study or for work projects? Arrangements might also work with area high schools, who might then feel obligation and loyalty to "feed" candidates into Pitt's

undergraduate programs. Such an arrangement could accomplish networking and establish economies of scale – a compelling argument to any financial office considering budget cuts.

The Team did not explore global and international trends in our own graduate schools. I know from personal experience that Pitt's full-time business program boasts a 30% international population, including strong contingents from Taiwan and India. Those students and programs might have knowledge and lessons to share – revealing not only active opportunities for global research for students, but also creating "feeder" programs into Pitt's own graduate schools in international study, comparative education, and social work.

While not absent from Team attention, input variables were not explored thoroughly and instead left to assumption. Internal investigation could have included quantitative profiles of what is proclaimed but necessarily proven to be a diverse student body. Empirically, Team members and planners might have observed faces of color throughout campus, but this did not mean that diverse needs are being served – only present, and compared to what standard? Although not critical to the workings of evaluation, those details could have made the outcome seem timely and compelling. For example, an outcome with spin might read, "In keeping with the needs and desires of a diverse student body that includes 28% international scholars, students will have many opportunities to gain diverse and global learning experiences". That statement still affords wallflowers the privilege to abstain, but grants collective recognition and active programming that seeks inclusion. Moreover, percentages and numbers on ethnicity and other student attributes could enable planners to fund programs according to budgetary "entitlements" and needs.

By omission, if internal scans do not reveal demographic percentages and figures on ethnicity, then the presumption is that the outcome pertains to homogenous groups that could

benefit from greater exposure to other cultures. Political ramifications aside, this outcome states valid intent that the majority of American-bred students should have opportunity to broaden their interests and awareness. Perhaps any polarizing caused by phrasing could be resolved by adding a clause on "inclusion".

A possible misuse of terminology might be to confuse global experience with "globalization". Could a student who orders products online from India, chats with foreign colleagues, and joins international organizations truly engage in global consciousness? If this outcome seeks to promote actual travel, language study, and study abroad – something beyond casual awareness that culture exists beyond national borders – then Student Affairs should revisit the wording of this outcome to ensure it is construed as such.

A semantic distinction is not to equate "learning experiences" with Student Learning Outcomes, a different set of academically focused outcomes in the domain of the College of Arts and Sciences and Provost's Office. It may be fine to cross-reference into other sets of outcomes, but Student Affairs' role remains to evaluate global experiences outside the classroom. The word "learning", while broad and generic, could generate confusion between sets of outcomes that are unique and defined within distinct operating units.

Another semantic criticism cites the definition of "many". What volume or percentage would qualify? Without fretting about definitions, the Team and managers could rely on calculated baseline measures, extrapolated growth in student interest, and commensurate funding and staffing to imbed a defensible and durable definition of "many".

Thus, next evaluative steps should relate to benchmarking that gives summative support and formative detail. Every international program should track how many students actually participate, and follow-up interviews should gather students' (and alumni!) impressions of their

experiences from overseas and diversity related programs. Mailing lists surely exist in alumni records, an area with which Student Affairs must collaborate to conduct evaluations with a longitudinal angle. Ideally, testimony would share how participants' lives were enhanced. Former students who valued these encounters could likely be encouraged to share positive feedback that would support continuation of global opportunities for next generations of scholar-citizens.

➔   The logic model reveals several significant lapses and gaps in the Team's evaluation.

4.2.3   Pitt Pathway (Student Baseline Outcome #11)

This sub-chapter analyzes a very good evaluation by the Team of an outcome deemed by Student Affairs leaders and University brass to have upmost strategic importance.

Outcome (Intended): "The entire Division will become involved in the Pitt Pathway Program; we will have evidence that the program is making a difference in the lives of students."

Assumptions.   Leadership considered the Pitt Pathway a strategic priority as a unifying theme that encapsulated mission towards "the best collegiate experience in the country". Planners conceived the Pathway as a prescribed route or customized journey through each student's academic career and social development, consisting of many "stations" or targets ranging from academic distributional requirements to community service to job interview training. An example of a station might be development of an e-portfolio – nicely tying together with an outcome analyzed previously – or a visit to Career Services to create a career plan, or attending workshops on how to craft a senior thesis. Determination of "making a difference in

the lives of students,"[39] would be determined through component program evaluations, since the Pathway consisted of many interwoven programs.

Environment. That the Team was unaware of similar programs at other campuses does not mean precedent and models did not exist. I feel that environmental analysis of this outcome was moot, as even negative testimony from literature and consultation would not have deterred leadership from making this outcome a primary plank of the assessment campaign. Management believed firmly in this outcome, and has been proven correct.

Inputs: *Student attributes*. This outcome stemmed partly from adventuresome planning and partly from frustration voiced by students in focus groups, RA feedback sessions, and annual surveys – amounting to a thorough needs audit. Student discontent centered around Panther Central, the student services headquarters. Originally centered in Career Services, Pitt Pathway evolved into a response to students' call for one-stop shopping and resolution of problems with registration, housing, maintenance, and services – whose organizational boundaries should be transparent to students. Moreover, Pitt Pathway was seen as a component of Student Affairs objectives to unify and improve communication within and beyond the division.

Team activity. The Pitt Pathway would most visibly become an enveloping theme later in the assessment campaign (after my departure). During my tenure, the Team's evaluation report to the Dean in February 2006 included an internal audit revealing that even departments that did not actively participate in Pitt Pathway still made appropriate referrals. As long as all staff would continue to market the program to students by enthusiastic interaction, Student Affairs could pass muster for this outcome. This could be an example of proper delegation to achieve an outcome – that is, if the Dean could tolerate a minimal approach where, throughout

---

[39] Source: University of Pittsburgh Student Affairs "OCC" (Outside the Classroom Curriculum) pamphlet, 2009.

110

the department, successful evaluation might depend upon closer collaboration with Career Services as the Pathway "nerve center".

In the second year of the campaign, this directive ballooned into a formal task force (which is a better example of departmental level assessment than of Team activity). The Assessment and Measurement Team collaborated with the Marketing and Communications Team on ways to attract students to the program vision and the meaning of various "stations".

Outputs: *Programs.* A major thrust was to expand the scope of Pitt Pathway beyond its origins in Career Services and to integrate the program throughout Student Affairs. Our Assessment Team began its analysis with an internal scan, yielding the following findings:

- Student Life claimed to connect the Pitt Pathway to program areas by including reference to the Pitt Pathway in web pages; publications such as the Student Life Brochure, the Student Organization Operations Manual and the SVO monthly Volunteer Opportunity Bulletin; signage; promotions, such as ads in the *Pitt News*. The Dean's presentation during New Student Orientation 2005 focused on the Pitt Pathway.

- Residence Life presented a symposium in October 2005 on "The Pitt Pathway and Residence Life's Web of Inclusion".

- At Student Health, students who participated in health education programs were "often referred to the Career Center to allow them to benefit from their services and promote the Pitt Pathway" ("Evaluation of Internships, Pitt Promise, Pitt Pathway", February 10, 2006, p. 5).

- Judicial Affairs reported "[not doing] too much with this, other than indicate that participation in Judicial Board could be a part of a student's Pitt Pathway. We ask students to find a community service opportunity that could turn in to a long-term participation or something that would complement their major whenever possible" (p. 5).

- Disability Services (DRS) reported that "staff threads the Pathway through regular case management sessions with students. Staff inquires students as to their status on the Pathway and documents responses… in case notes" (p. 5). This seemed a sensible, customized approach that applied a broad concept to individual needs.

DRS created a bulletin board in the reception area dedicated to the Pathway. The office arranged for the Program Manager of Pitt Pathway to speak to staff about the Pathway and how to integrate it into students' active consciousness – a training approach that might appeal most to staff who have experience and belief in counseling and student development theories. In that approach DRS offered a valuable lesson on how to achieve buy-in for new or unfamiliar programs by tapping into familiar practices and mindset. Keeping with trends on technological progress and connectivity, DRS planned to reference and link the Pathway from its own web page. The office suggested that other departmental and unit websites follow suit.

- International Services (OIS) claimed to have no dedicated interface with Pitt Pathway other than cooperation with Career Services on the PASSPORT (Preparing and Assisting Students and Scholars at Pitt through Orientation, Resources, and Training) session, "if that should be viewed as part of Pitt Pathway" ("Evaluation of Internships, Pitt Promise, Pitt Pathway", February 10, 2006, p. 5). In our report, the Team opined that PASSPORT should indeed represent a "station", on grounds that units should not downplay the efficacy and participation levels of existing programs that serve to further an outcome by any magnitude.

- Career Services provided the administrative oversight for the Pitt Pathway and sponsored its Program Manager position. Career Counselors and Career Consultants regularly promoted the Pitt Pathway and related activities in individual appointments and in group meetings with students. By design and intent, all programs offered by Career Services fell

within the scope of Pitt Pathway. The Emerging Leaders program devoted an entire session to Pitt Pathway and related activities, enlisting Career Services Counseling and Consulting staff to facilitate that session.

- Other than support in training sessions for Emerging Leaders, the Counseling Center did not claim to have specific programs for Pitt Pathway, although the unit's developmental approach to therapy encompassed tenets of the program. For specific work on Pitt Pathway, this office usually referred clients to Career Services. This seemed an example of collaboration with areas better equipped with given solutions, than deflecting a new directive. Attention to this outcome would possibly encourage the Counseling Center to adopt an active role in Pitt Pathway, perhaps by marketing and de-stigmatizing students' needs for counseling in certain circumstances and by thus depicting the Center as an important "stop" along many students' voyage.

➔ The most notable advancement has been the "Outside the Classroom Curriculum" (OCC). This program was implemented after my tenure, and out of scope for detailed analysis, except as illustration of how an outcome informs later stages – how information flows back down the slope of the pyramid logic model – to reconfigure assumptions, frame inputs, interpret environmental factors, and suggest formative programming.

Outcome (Actual). *Results*. Without cohesive marketing and unified buy-in, this program did not congeal at first, leaving students confused over its intent. As a fellow student, I thought the Pathway concept was an attempt to market and structure progress that should have been intuitive. Buy-in has since gained momentum as Student Affairs leadership has pushed this initiative towards the top of their agenda. Participation would become a key element of follow-up as the program gained credibility and publicity.

Outcomes may unfold over time, and this one has been no exception. In the long term, Pitt graduates in a longitudinal study may inform evaluators that the Pitt Pathway was a pivotal influence along their career path. In the shorter term, a 20-person task force refashioned the Pathway into the Outside the Classroom Curriculum (OCC), launched in Fall 2008. While the program's timing – after my involvement -- technically places investigation of program particulars beyond scope of this study, I would be remiss in not explaining how to work results into re-configurations of my logic model. A feature of the logic model is how it can be updated by inserting details into codes and classifications. For this outcome, pertinent benchmarks include 1) current registration of 6,057 Pitt students, 2) the forthcoming graduation of the first student with OCC distinction, and 3) positive qualitative feedback from students, such as, "The OCC has exposed me to many things that I normally would not have been exposed to" (Thrasher, 2009, p. 1).

Figure 20, ahead on page 117, shows how this outcome fits into the logic model's depiction of mission and vision. I feel this outcome has been the Student Affairs assessment campaign's most compelling statement both of strategic planning and brand marketing.

Meta-evaluation.

Figure 19   Logic model depiction of evaluation of Pitt Pathway outcome
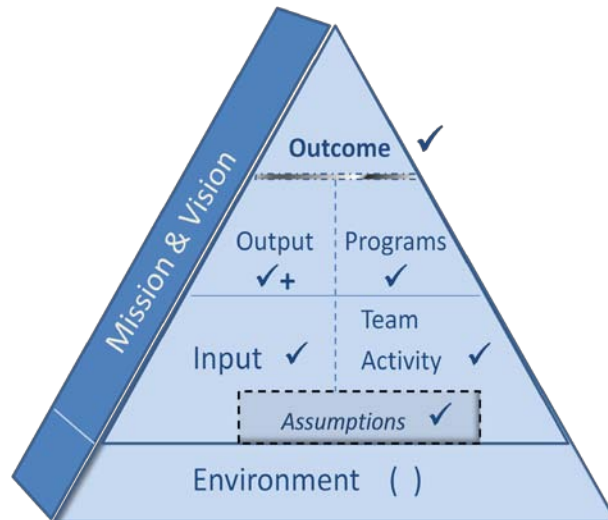
The Team performed well in seeking evidence for an outcome tied deliberately to mission and vision. Time will determine the ongoing success of this outcome, as future surveys should measure alumni perceptions of value. The logic model signals long-range activities of benchmarking from qualitative feedback, possibly transforming into quantitative approval ratings.

I would rephrase outcome text from "having evidence" to something less outcome-driven and more process-based, such as "seeking evidence". Demanding a presence of evidence both presumes causality and that the outcome cannot be true nor achieved without evidence. Patton (1990) and Weiss (1997) warn against ascribing causality yet also against ignoring unintended outcomes, which may be favorable. Another angle would be to disprove alternative explanations (Weiss, 1997; Patton, 1990), for example by surveying non-participants, if that were not such a difficult chore.

Detractors might claim that this outcome reeked of politics, in assembling alliances and task forces between areas of separate jurisdiction, and in attempting to extend a program beyond

its original confines. The point is that students likely knew the program was contrived, but hopefully for the right reasons, with their development and success in mind. Although intentions were good, effort might have been better focused on direction within departments and units rather than fitting into a grand scheme.

Proponents would say that this program, in serving mission and vision, remained planted within political boundaries. Recent coverage in university publications affirms that this outcome remains a focal, visible, and assertive, and that programs related to Pathway / OCC have gained remarkable momentum. I find this impact both surprising and favorable – definitely a feather in the cap of Student Affairs leadership. The Team should believe that our efforts helped inform and guide these changes. Already Student Affairs has seen positive feedback, growth in participation, and benchmark measures. To reiterate, participation has grown to more than 6,000 students as of Spring 2009, and the first student to graduate with OCC distinction will do so in April 2009 (Thrasher, 2009).

In closing on this outcome, I suggest modification of my pyramid logic model to allow individual and detailed assessment of each of nine OCC metrics: Leadership development, sense of self, career preparation, university participation, communication skills, respect for diversity, healthy lifestyle, service to others, and appreciation for the arts. Driscoll and de Noriega (2006) view assessment through a trifocal lens of mission, vision, and core values. While most outcomes have been framed only by mission and vision, I equate the nine principles that comprise OCC with "core values". Future assessment of the Pitt Pathway / OCC should discern each core value, as depicted in the following figure.

Figure 20   Figure 1 rotated: Logic model depiction of core values that frame Pathway assessment

OCC / Pathway Vision = "Educating the Whole Student"
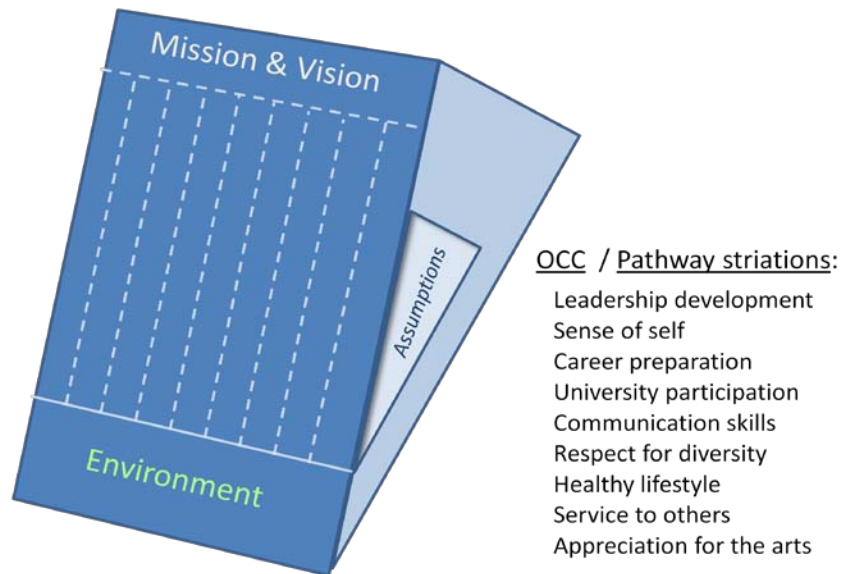
The usual "front face" of the model, in light blue, would look the same as previous renderings, except that "assumptions" are highlighted for prime linkage to core values. Accompanying narrative analysis might not be based upon one logic model, but upon nine separate models that add up to Student Affairs vision of "educating the whole student". Each "programs" block would have different content. For example, assessment of leadership development will include evaluation of the Emerging Leaders Program, of a new leadership LLC in a Pitt residence hall (White, 2008; Brooks, 2008), and of OCC "transcripts" documenting student participation in various checkpoint programs and culminating in a "green cord of distinction to wear at commencement" (White, 2008, p.3). Evaluation of all core value facets should include follow-up with graduates on how "their experiences have transformed their lives and made them the types of people a company wants to hire or that a school wants to admit" (K. Humphrey, quoted by White, 2008, p.3).

➔ In sum, this outcome represents Pitt Student Affairs management's finest triumph in instilling a culture of assessment, a good job by the Team to investigate an outcome, and a stellar illustration of how a logic model can be used to analyze evaluation process. It is also a positive note upon which to wrap up this portion of analysis and proceed to meta-evaluation of the whole campaign.

4.3 FINDINGS / META-EVALUATION

How does this add up? One possible way to determine program success would be to tally the number of outcomes rated positively. The matrix score-card on the next page summarizes and rates the Team's activities and the Student Affairs division's campaign according to the rubric below:

Table 6   **Key to ratings matrix on overall campaign**

Symbol	Meaning	Point value
✓	Element has been analyzed adequately.	1
✓+	Team did an exemplary job in this task.	1 + ½
(✓)	Evaluated but unresolved or conditional.	½
( )	Not resolved or determinate.	Ø
-	Element or task was not resolved.	Ø

In Table 7, the second column tells the outcome number as listed in the Pitt Student Affairs Student Baselines Outcomes Document (also given in Appendix A). The "overall" column, next to last, gives my overall rating of the Team's performance on that outcome. These are subjective judgments, what I perceive to be a qualitative "average" of all marks for that outcome. Where the majority of elements are complete, the evaluation gets a check "✓". Missing

118

important factors deducts an otherwise complete evaluation to "(✔)". A few outcomes were not completely evaluated in the time frame of this study, and receive a rating of indeterminate "( )".

The final column gives scores that tally ratings of each logic model element (i.e., "input", "environment"…), according to the scoring rubric shown in Table 6. The final two columns thus represent different ways of looking at overall performance on a given outcome. Both bases may be used to judge the Team's overall performance, depending on decision makers' preferred variables. Using the same metric in table 6 above, the bottom row of Table 7 measures quantitatively the Team's performance on key factors ("input", "output", etc.) across outcomes.

Table 7   **Assessment Team "score-card" for Outcomes, 2005 – 07 campaigns**

Outcome (intended)	Outcome # / sub-ch #	Report release date	Logic model elements status						Overall	Total
			E	I	Activity & programs	Assump -tions	Output / results	Outcome		
E-portfolios	7 / (4.2.1)	Dec. 9, 2005	✓	--	✓	✓	✓	✓+	✓	5.5
Mentoring	12 / (B.1)	Jan. 31, 2006	✓+	--	✓	✓	✓	( )	✓	4.5
Internships	8 / (.B.2)	Feb. 10, 2006	✓	✓	✓	✓	✓	✓	✓	6
Pitt Pathway (OCC)	11 (4.2.3)	Feb. 10, 2006	( )	✓	✓	✓	✓+	✓	✓	5.5
Pitt Promise	5 / (B.3)	Feb. 10, 2006	--	--	✓	✓	✓	( )	--	3
Listening	3 / (B.4)	Mar. 22, 2006	( )	✓+	✓	✓	✓+	( )	✓	5
Excellent experience	1 / (B.5)	Mar. 22, 2006	--	✓	( )	✓	✓	(✓)	✓	3.5
Student Affairs performance (> average)	2 / (B.6)	May 18, 2006	--	✓	✓	✓	✓	(✓)	✓	4.5
Global experience	4 / (4.2.2)	May 18, 2006	--	--	( )	✓	✓	(✓)	(✓)	2.5
Cut programs	6 / (B.7)	May 18, 2006	( )	✓	✓	( )	( )	(✓)	(✓)	2.5
Parents conduit	14 / (B.8)	May 19, 2006	✓+	✓	✓	✓	✓	(✓)	✓	6
Faculty liaison & tools	13 / (B.9)	Feb. 6, 2007	✓	✓	✓+	✓	✓	(✓)	✓	6
Alcohol violations	9 / (B.10)	?	--	--	( )	✓	--	( )	--	1
Conduct Code	10 / (B.11)	?	✓	--	( )	✓	--	( )	--	2
TOTAL			7	8.5	10.5	13	12	6.5	10	

One way to conduct overall meta-evaluation is to add up overall ratings from the "Overall" column. Check marks add up to success in this case, by a 9 to 3 margin, plus two unresolved, for a score of 10 out of 14. The "Total" column then supplements this result by offering individual totals; for example, the Team did relatively well evaluating Internships (6 points), Pitt Pathway (5.5 points), e-portfolios (5.5 points), parents' conduit (6 points), and faculty tools (6 points). Weaknesses show on evaluating global experiences (2.5) and summative evaluations (2.5). (Alcohol and Conduct Code violations require more data to meta-evaluate performance.) This seems an accurate and meaningful way to meta-evaluate the campaign according to outcome resolution; a manager can glance at these scores and determine where the Team conducted thorough evaluations.

Another way to meta-evaluate this campaign is on basis of elements. If a manager scans the bottom row, she will see that the Team did a great job with assumptions (13 out of 14 points) and outputs (12), but the Team needs to work harder on gathering information on environmental factors (7 of 14) and inputs (8.5). [Note: 6.5 in the "outcomes" column does not signal weak resolution of outcomes, but that several outcomes have not yet been resolved. This score will likely increase over time.]

Accordingly there could be a "score pyramid", too. The figure below shows ratings per key element, according to the same scoring rubric in Table 6. Corresponding percentages are calculated by using 14 (the total number of outcomes) as the denominator.

Figure 21   Team performance scorecard based on logic model pyramid
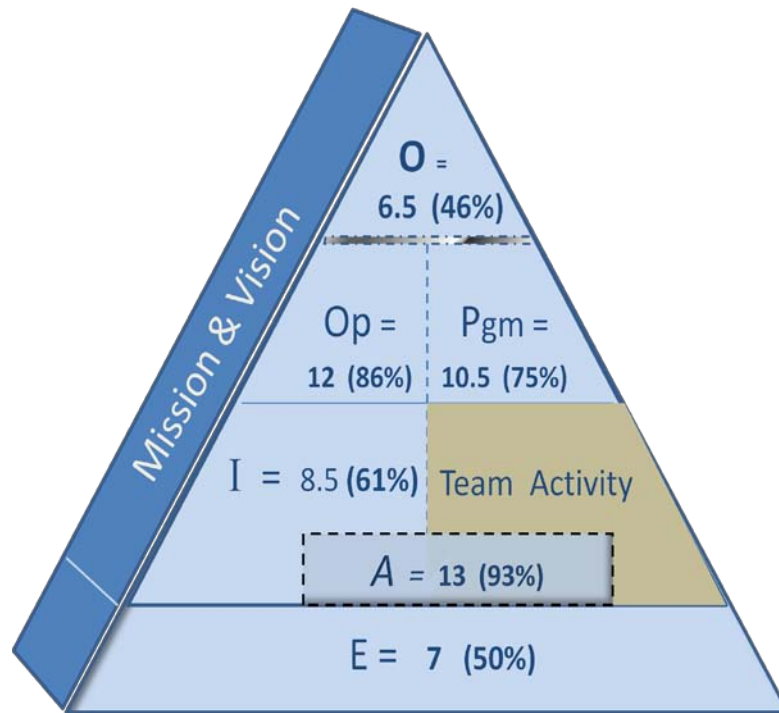
The "Team activity" slot is blocked out, except for "assumptions", because Team activity is what this diagram evaluates overall. Scores and percentages may be misleading, particularly for "Outcomes". The 46% does not mean that the Team did poorly in satisfying only six-plus outcomes, but rather that this phase of the assessment cycle is 46% complete at this snapshot moment. While the "outcomes" column in Table 7 shows many parentheticals, lack of resolution illustrates the nature of outcomes: that they might unfold as time goes on. The Team did a good job looking at nearly every one and giving an estimated status. Only "alcohol violations" and "conduct code violations" remained uninvestigated during the time frame of this study. The logic model points out those gaps.

I interpret that other percentages do reflect Team performance on key elements. Percentages indicate that the Team should conduct more thorough environmental scans. The logic model is useful in pointing out missing pieces and providing status information, so that evaluators know where to focus attention next.

Scanning down columns in Table 7 and looking at percentages in Figure 21, I observe that the Team performed best (that is, the greatest concentration of check-marks) with "assumptions". These scores offer support that the Team handled assumptions well[40]. I believe this observation points out the value of a team, with members who cross-check facts and challenge assumptions. A single practitioner might not have similar resources to test assumptions against actual events. This finding supports management's subsequent decision for the Team to take on an advisory role that capitalizes on their collective strength in formulating and revising assumptions that drive activity on unit levels – where greater accountability now resides.

Table 7 indicates that the Team also did a commendable job locating programs, both active and precedent. The Team was not as vigilant in gathering "input" data. One could argue that this "input" element is the most critical, as it represents what students believe they need and want. Less than thorough attention to input variables compromises internal validity, the link between inputs and outputs (Weiss, 1999, p. 241). Needs audits and other methods to emphasize inputs should be of foremost concern in Student Affairs. I feel that the Team would be wise to conduct thorough "needs audits" on every outcome revisited in upcoming seasons of evaluation – switching from the "Assumption #2" (p. 65) approach used previously of assuming such audit has been conducted to the "Assumption #1" approach of summative accountability.

➔   By the measures outlined here, the Team did well. However, **a key finding is that before enlisting this logic model to analyze the Team's performance, I did not know for sure what the Team's relative strengths and weaknesses were.**

To summarize: A logic model can aid and enhance this process by introducing inventive ways of looking at issues and conjuring possible resolutions. I conjecture that pyramids of

---

[40] An alternative explanation is that my logic model enabled me to capture assumptions effectively. I feel that both interpretations support my arguments for using logic models to analyze and guide evaluations.

ratings, and my whole process of meta-evaluation, would come as a surprise to management and to the Team, and probably a pleasant surprise. Scores indicate that the Team performed well, especially as rookies. I also believe that, had we known of logic models and implemented one as standard procedure, Table 7 would contain no minus signs or parentheticals other than for outcomes not yet resolved. The logic model would have alerted us to missing pieces, which is one function of a logic model. How might I test this supposition? The best way would be to encourage the Team to use a logic model, and then to see how robust the next iteration of Table 7 turns out.

Analysis proceeds to speculations how the Team might enlist or customize this same model or another one to guide future phases of evaluations. In response to my second research question, I extrapolate that if a logic model can reactively explain what happened before, then perhaps a logic model can also proactively inform and guide next steps: future phases of benchmarking, possible re-formulation of outcomes and parallel assumptions, and gathering departmental perceptions about the campaign. Extrapolating a step further, I also hope that Student Affairs practitioners on other campuses may draw lessons and parallels pertinent to their own programs by recognizing how a logic model can signal, identify, and possibly prompt success factors, lapses, and consequences.

5.0   CONCLUSION


Scriven (2003) proclaims:

> At the most basic level, evaluation is a survival-developed brain function of supreme importance. In its simplest forms, it is close to the perceptual level, the near-instinctive process, sometimes inherited, sometimes learnt, which provides the kind of knowledge that links perception with action. One step up, and we're at primitive forms of real evaluation, which leads us on occasions to flee from the mammoth and on others to attack it, after we review its state and our resources. In its most complex form, evaluation elevates the simple process of instant or near-instant appraisal to the complex and systematic – and more reliable – investigation of elaborate systems, in this process becoming, of necessity, a complex discipline itself (p. 28).

In prose less heraldic, I conclude that the Team portrayed in this study slew nor fled no mammoths, but we did help make some college kids' experiences more enjoyable. In the process, Pitt conducted an assessment campaign that boasted key attributes of being "critical and supportive. This phrase is crucial. Without a critical assessment there is no growth[;] without support there is no acceptance" (Eisner, 1990, pp. 116-17).

As I stated in the Summary (chapter 4.3), a "sum" of positive evaluations balanced against outcomes not achieved can indicate the success of a campaign. I have taken a formative stance towards meta-evaluation and presumed summative merit, whereby quantifiable measures signaled by a logic model identify shortcomings and alert planners to improve processes that are inherently purposeful. A good logic model reconciles lapses and fills gaps by guiding the process and diagramming knowledge flows (represented by the green arrow in the figure below)

from outcomes, filtering back down slope to assumptions for next phases of evaluation – comprising a complete evaluation cycle for that outcome or program.

Figure 22   Logic model depiction of "impact": Outcome-based knowledge reframes assumptions
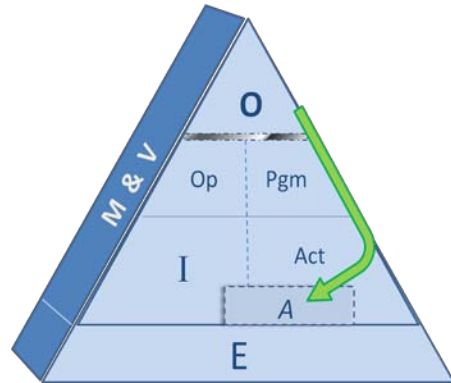
The result is that in the next phase of evaluation, while "environment" remains more or less static[41], "assumptions" likely change; inputs have detailed layers, and fresh outputs and programs utilize benchmarks. From quantitative and qualitative support, outcomes can in turn be restated to include specific targets and goals, or modified by supporting directives. For example, the next iteration of Student Baseline Outcome #11 might state, "The entire Division's involvement in the Outside the Classroom Curriculum will gather evidence that the program is making a difference in the lives of students," and sub-texts – one for each of nine core values – might declare, i.e., "Leadership programs will encourage a majority of students ($> 50\%$) to serve in a committee, club, or residence hall leadership position". This is but one example of how assessment campaigns can proceed from singular statements of mission and vision to richer statements of data support.

---

[41] By "static" I mean that planners and evaluators cannot change its character. Consider recent trends in American and global equity markets. Environment has changed, but not because of anyone's programming. Meanwhile, an evaluator's assumptions _must_ change to reflect expectations from different economic conditions.

These adjustments to thought and process may signal that if a logic model works retroactively, then it may be enlisted proactively. I hope that my analysis has opened discussion on how this Team or any evaluation team could use a logic model to proceed into subsequent phases of evaluation and assessment. Before proceeding so boldly, the evaluator and researcher should consider limitations of this study.

5.1   LIMITATIONS AND ALTERNATIVES

My criticism of methods described here is that even with a rubric, ratings could be subjective. Ratings similar to a "grade" might encourage evaluators to seek a performance basis over truths about what was happening. Another criticism of collecting scores is the "snapshot" quality, despite the fact that each of these outcomes represents an ongoing process that may unfold into greater (or lesser) measures of success. A flexible mapping of variables into classifications does not guarantee that every column in Table 7 must have data. Certain variables may overlap or may be better described as, for example, environmental than programmatic. Still, the scorecard is one defensible way to measure a campaign's success.

If a scorecard matrix is used, I advise deferring to decision makers' judgments how a score denotes success. A corollary might assign weights to outcomes deemed most critical. For example, should Pitt students' perceptions of an "excellent experience" count double? Should "undecided" resolutions count for partial credit? Instead of check marks, an analyst might assign values (i.e., 1–5 or 1–10 ratings) for the Team's performance on each item. In formal assessments, ratings require a formal rubric in advance, explaining what elements must be present to "earn" a specific numerical rating or "passing" grade. Here I used a simple one-point

scale for "checks" and half-credit gained for plusses and deducted for minus signs and parentheses, according to the key in Table 6. Alternative scoring rubrics might result in cases where 4 of 14 (or any reasonable number) outcomes coming up favorably might not necessarily constitute a failed campaign. Perhaps those four were of critical importance and lengthy, exclusive duration.

Another purely qualitative way to judge success is to regard this case study as an evolving and renewed campaign, one that has progressed from its incipient to novice phases. Planners can determine its success by momentum, positive energy, and belief in the program and its intended outcomes. Student Affairs departments began with a rudimentary understanding of the assessment process. One illustration of this claim is early (circa November 2005) documents that state outcomes and their measurement only in terms of mission and vision. That in itself is admirable but should now be modified with concrete, measurable goals. For example, the Residence Life and Student Health versions of Student Baseline Outcomes documents referred generally to focus groups, comment and suggestion boxes, professional development, newsletters, open door policies, "inclusion", and improved marketing as utilities and illustrations of commitment to desired outcomes. These were noble, albeit bland statements of tasks. Statements of intent, while well intentioned, lacked necessary detail and specificity, betraying a naïveté towards the process. The Team learned a lot since then, and throughout the campaign reports took on a confidently exploratory character.

While the Team did not display actual aversion to quantitative details, only evaluation reports on "listening", "excellence", and "above average performance" contained numerical measures conducive to continued benchmarking. While those examples demonstrate that practitioners had learned to trust the process and to articulate measures in definable, measurable

128

terms, often the Team deferred to "easier" narrative description for other outcomes. The better news is that quantitative back-up has been readily available from unit-level reports, especially analysis of the annual Quality of Life survey.

I feel that quantifying results does not compromise a qualitative researcher's preference that each process should be treated holistically and independently on its own merits, even when practitioners may observe overlaps between processes in terms of actors, resources, physical space, timing, and other qualifiers. Eisner (1990) assures that in qualitative research, "numbers are okay"[42] (p. 186).

Another limitation involves programming scope. Were evaluation samples representative? Weiss (1997) reminds the researcher that control groups and null sets are not equivalent. "Any experiment does not pit a program versus nothing; it compares a program to the usual run of experience" (p. 220). The "usual run" might deliver less than intended results, and "it could turn out that the people most in need are not those helped most by the program" (p. 218). Often I have seen lists of participants include only, as I called them, the "usual suspects" who had already subscribed to Residence Life directives. Meanwhile reclusive students were neglected by no one's fault. Despite Weiss' claim that evaluation uses control groups to seek whether programs provide benefits beyond those of common experience, a less than common experience often goes unexamined. I see this duty to serve an under-the-radar student population as a nearly irresolvable dilemma in Student Affairs programming. Many students do not need programming; others shun it, and others cherish privacy such that J.D. Salinger would sooner grant an interview. A student sample usually does not include the unreachables who are immune

---

[42] Eisner is so emphatic about this principle that he says it twice in the text: "Numbers are okay!" (p. 186). Patton (1990) offers a more resounding endorsement: "Numbers convey a sense of precision and accuracy…" (p. 479).

to campus programming. Patton (1990) suggests that negative and deviant cases add interest to analysis (p. 463), but I note the difficulty for Student Affairs to please wallflowers and non-respondents who indignantly ask for removal of their names from distribution lists[44]. Because the human world is not ordered and the researcher not omniscient, Patton warns that perfect patterns might be greeted skeptically (p. 464) in what I call an evidentiary paradox. Patton takes a somewhat softer stance than Weiss that alternatives need not necessarily be disproven, and I suspect he implies that the chosen finding is best. In the Student Affairs setting, I offer experience that programming can be enhanced by collective perspectives that facilitate students' searches for meanings.

This study has its basis planted in experimental research and programming, which introduce new variables but do not escape criticism. Weiss (1997) cites the remarks of Mary Lee Smith (1994) that "experiments decontextualize action and distort effects, shift authority from practitioners to evaluators, and attempt to control… conditions that results do not generalize" (pp. 229, 232). In defense of regular as opposed to impromptu experimentation, Weiss claims that a "repetition of results may spur a standard of firm knowledge" (p. 235). This replication of results can prompt incremental improvements in programming, a healthier and steadier approach than implementing urgent policy adjustments upon premature findings. Weiss (1997) acknowledges rejoinders that a by-product of repeated evaluations has been inconsistent and sometimes conflicting findings about the same program. Even a good logic model cannot rescue flawed methodology.

Others take a purely means to an end approach. In his critique of assessments of comparative education systems, Husen (1997) offers valuable advice to the evaluator: "What

---

[44] In a nasty e-mail reply to a survey invitation, some brat once threatened legal recourse for spamming. Evaluators develop "bulldog skin" (GBV, 1997).

they did or did not do should not be judged by the hindsight knowledge and eventual wisdom of today" (p. 43). I concur that strengths and weaknesses matter less what they were than whether they have been revealed to alert policy-makers. My logic model has demonstrated such completions and gaps in evaluation process.

Leadership and personality. Leadership may be a critical force in a campaign's success. One way to test this theory would be to locate successful campaigns (or unsuccessful ones) led by various personalities. One could inquire, what happens if assessment is launched because a charismatic leader insists that it will work? Hype may disguise a lack of quality. If the campaign does work, what happens if that person leaves? Has her legacy been enough to foster an eternally rosy disposition towards assessment? Enthusiasm may be contagious, even in delicate areas of buy-in. But such enthusiasm can seldom disguise a meritless or unstructured campaign. Conversely, why could a dull individual not spearhead a well-planned campaign? Focusing on a bland personality may dispatch an improper and inaccurate perception that useful action, like its program engineer, lacks punch.

Documents offer evidence that this campaign did not happen only because a charismatic leader wanted it. Rather, enthusiasm inspired a team to get the job done. Meeting minutes contain the following quotes from an inspirational pep talk that commenced the first Team meeting ("Assessment Team Meeting Minutes", September 8, 2005, p. 1):

> *"The Charge!" The Assessment Team was the idea that inspired all the other new committees in Student Affairs.*
> *A list of [outcomes] guides our activity. Of these, some will be accomplished; all will have progress.*
> *Make it plain, not pretty.*
> *Status quo is not acceptable.*
> *We aspire to a culture of assessment. Our team will lead in creation of that culture within our operations.*
> *The team approach prevents personal ownership by departments and individuals.*

131

[Leadership] professes opinions how it should be done, but grants us the autonomy and judgment to forge best assessment practices.
Our efforts and dynamics represent a mirror; we cannot deny what we see.
Internally, our primary outcome is to promote a cohesive, unified division, instead of several units.
Looking outward, assessment will serve our vision "to provide Pitt students with the best collegiate experience in the country".

Who could resist "the Charge"? A critic might inquire whether these sentiments betray an over-emphasis on personality. Without generalizing that a dynamic personality must always be the case, or conversely that a boring leader could not have adequately spearheaded an efficient campaign, I claim that this particular leader's personality and enthusiasm positively contributed to the evaluation process, to the achievement of outcomes, and to a "culture of assessment" taking hold.

To extrapolate, not to generalize.   A topic related to limitations is generalization. Even if a result or technique is striking and significant, how far does its applicability carry? Weiss (1997) says that a single program is "prisoner to its own setting" (p. 236). At the same time, that single program adds piecemeal wisdom to a growing and evolving consensus.

What made this dissertation interesting for me was to apply a model retroactively. Logic modeling offers a contrast to what actually happened. All our Team knew about was "standard" techniques related to evaluation – internal and external scans, surveys and focus groups – but we did not have a method or plan to organize our data collections or to envision the whole process holistically. A logic model can congeal concurrent functions into a unified, cohesive campaign.

While I hope that other researchers and evaluators are tempted to test and perchance use my logic model, I recognize a warning to avoid leaping into generalizations about applicability of this model anywhere, even in the same context as the case study. A main plank of my argument is that functional logic models might replicate into useful standards and precedent for

generations of institutions that adjust to changing parameters. Goodstein, Nolan & Pfeiffer (1993) caution about distinctions between proactive planning and the dangers of "forecasting", which may result in unrealistic expectations and reformulation of plans. While a critic would unlikely contend that my logic model represents a failed application, one could assert that hundreds of other models and approaches could work better here or anywhere else. A consideration I find problematic is whether a simpler linear model, Astin's (1991) I-E-O model in particular, could have arrived upon the same results, or sufficient findings. While a follow-up study could test that hypothesis, I stand behind my original model being both more durable and more interesting, especially when researchers and evaluators consider weaknesses of Astin's model in 1) not distinguishing outputs from outcomes and 2) leaving a modest arrow to denote assumptions and causality clearly defined and situated in my model. I prefer how my model channels logic and thought upward, to the apex where outcomes belong. I think that my model's inventive quality could appeal to other researchers and evaluators who depart from thinking along linear paths of theory and practice.

If a researcher knows never to generalize, then how do we borrow potentially useful models and ideas? Patton (1990) joins other theorists in progressing from generalizations to *extrapolations*. "Generalizations decay…, valid only as history", and always context free (Patton, 1990, p. 486, citing Cronbach 1975; Guba & Lincoln, 1981, p. 62). Crediting Cronbach, et al (1980) for this insight, Patton elaborates that effective evaluation should strike "a middle ground stance to permit extrapolation, which connotes that one has gone beyond the narrow confines of data to think about other applications and findings" (p. 489). This caution tempers claims of Stake (1978) that to generalize borders on idiocy. A useful result may come from "modest speculations" (Patton, 1990) under similar but not identical conditions.

133

I posit that a requirement of this study's extrapolation is a healthy organizational structure where some type of ordered campaign could take root – with understanding that these techniques offer no infallible panacea. Popham (2008) reminds the idealist, "Don't let pursuit of the instructionally perfect prevent you from reaping the rewards of the instructionally possible" (p. ix). I have seen that healthy teams involve compatibility, a human element beyond professional commitment. I would be remiss in not mentioning the role of various personalities who shaped this assessment campaign, which would not have been possible without the mutual respect shared among Student Affairs practitioners. I hope that I have been able to convey in these pages my fondness and appreciation for every member of our Team. It was an honor to serve as Co-Chair.

5.2   NEXT STEPS

One heartening conclusion derived from lengthy analysis is that assessment has worked wonders already for Pitt Student Affairs. New ideas presented in this study can encourage continued and perhaps greater success.

Evidence comes from managers' testimony that unit leaders currently conduct evaluation and assessment and continue to believe in it. Support staff and those entrusted with process and monitoring have not mutinied or sat idle. Their attitude remains positive and open. They enlist advice from the Team, still active in a consultative role. Learning has trumped resentment and confusion over the meaning of evaluation. Formerly feared, the process has become part of standard operations and expectations. One weakness of these observations is that they amount to

hearsay. More compelling evidence could come from surveys[45] of Student Affairs staff and leaders on their opinions about the success of the assessment campaign and how they feel about changes in team roles. Analysis could reveal whether the culture of assessment has been ingrained, or conversely whether Directors feel burdened with "extra" tasks. Survey responses might offer speculation, preliminary results, and first-hand data pertaining to linkages between outcomes and policy change.

Such steps are part of a living, evolving logic model, since future observation and achievement of intended outcomes could reveal the wisdom and success of approaches selected among several candidate approaches to purposeful assessment. Changes in logic model content and flow – particularly in "assumptions" and administrative "inputs" boxes – depict shifts in assessment process and theory. My analysis has not attempted to resolve specific issues or meta-evaluate management styles and adaptations, but has showed where evaluation components fit into a comprehensive framework.

I return to a realization that meta-analysis has limits, along Patton's (1990) advice to progress from generalizations to extrapolations. If the logic model has worked in this study, it may work in other contexts. I have tested whether using a logic model can reveal how needs link to perceived outcomes, preserving internal validity of one whole campaign. The logic model rests on assumptions that college or department administration and a designated team can derive and reconfigure outcomes from accurate, true-faith depictions of student input and environmental attributes – from evidence. As my tests have yielded positive results, then Student Affairs practitioners at the featured campus and other ones might find reason to use this model or a similar one for their own evaluations and assessment.

---

[45] IRB would not allow me to conduct such a survey without formal sign-offs from every potential respondent, rendering the survey impractical.

Then, this paper's conclusion should offer modest speculations upon simulated assessment, as if Pitt Student Affairs had hired me to tell them what to do next. My recommendations would include the following:

1) Use a logic model – if not this one, then one designed by a task force consensus to structure, organize, and track the ongoing campaign. WKKF (2004) reminds potential users that "there is no best logic model…; try several on for size" (p. 13). The authors continue,

> Like anything else, it takes practice to use logic models as effective program tools. We learn through trial and error to find what works best for what program. Don't hesitate to experiment with program logic model design to determine what works best for your program. And don't be concerned if your model doesn't look like… case examples (p. 13).

I have experienced that designing and customizing a logic model stirs creativity and may even add to the researcher's credibility, as someone not content with standard or familiar models.

How could Student Affairs put into institutional usage logic modeling as an assessment guide? I would recommend piloting a logic model in the next phase of evaluation upon one outcome. Analysis could progress layer by layer, with different status codes as demonstrated in the following figure:

Figure 23   **Process logic model with status codes per key element**
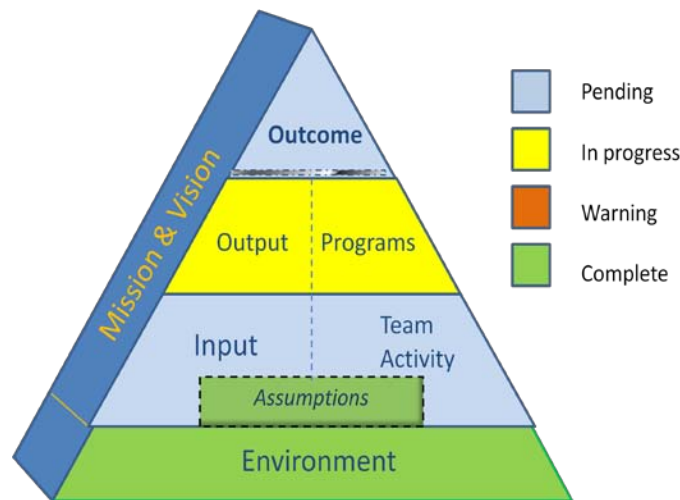
In this example, the environmental scan has been completed, and assumptions are formulated. An internal scan of input factors, such as student attributes, remains pending. Other internal scans are looking at existing programs; if problems are detected, or formative improvements called for, the evaluator may elect to color the "program" block red. A summative judgment might become a darker shade of red. If the director were expecting results urgently, the evaluator might code the "output" block red. Completion dates and target dates could also be useful information inserted into slots or a separate timeline graphic. I would recommend that evaluators experiment with various codes and keys to determine what audiences of stakeholders like best or relate to most readily. All of these functions could be supported in PowerPoint and linked to other utilities, such as an Access database or Excel spreadsheet. Diagrams could also be hand drawn on white boards. Planners could consider a permanent white board in a key strategic room, the "Assessment Room" in some under-utilized space (if there is such a place). The Technology Team could parallel this master planning board with an interactive / wiki website. The key is access for teams of users who know where to locate and share information on current and archived evaluation projects.

2) Determine causality for each outcome at reasonable intervals by using benchmarks and follow-up interviews and surveys. However, "causality" should not be a formal element in a logic model or slot in a matrix, but should be implied by logical flows of knowledge and assumptions.

3) Re-formulate policy and programming based on findings. If no causality can be determined, a) consider summative evaluation of the program, or b) identify different or missing

elements of the logic model for that outcome – then try evaluating the program again with fresh assumptions.

4) Communicate findings, policy decisions, and impact – all with evidence – to stakeholders. I summarize analysis as the evaluator's commitment to explain the quality of data, allowing a decision maker to implement sound policy based on measurable evidence. Some of that measurable evidence provides a baseline for further benchmarking, fostered by logic modeling.

➔ I conclude that an evaluator's up-to-date toolkit includes logic modeling. This study demonstrates how one particular model can prove useful, flexible, and enduring in a formal assessment campaign. In the process of using this model to improve evaluations that in turn lead to formative improvements and summative justifications of programs, every Student Affairs department may aspire to a compelling objective of "the best collegiate experience in the country" (University of Pittsburgh Student Affairs, Vision Statement, 2005).

STUDENT BASELINE OUTCOMES (REVISED JANUARY 2006)

1.   Students will rate their Pitt experience excellent.

2.   Students will rate all Division of Student Affairs programs and services above average.

3.   Students will feel that we have listened to them.

4.   Students will have many opportunities to gain diverse learning experiences.

5.   Students will be constantly reminded of the Pitt Promise to Civility.

6.   All programs that do not adequately meet the needs of students will be modified or eradicated.

7.   Student electronic portfolios will be investigated to determine feasibility.

8.   Every Student Affairs unit will create non-paid internships to create additional learning experiences.

9.   The number of alcohol incidences will decrease.

10.   The number of repeat offenders in violation of the Student Code of Conduct will decrease.

11.   The entire Division will become involved in the Pitt Pathway Program; we will have evidence that the program is making a difference in the lives of students.

12.   A Mentoring Community Project will be created and piloted this year and fully implemented in 2007.

13.   Faculty will be provided with additional tools to better understand and assist students.

14.   A conduit of information for parents will be created and utilized.

APPENDIX B

ANALYSIS OF TEAM'S REMAINING EVALUATIONS OF OUTCOMES

B.1   MENTORING COMMUNITIES (#12)

Outcome (Intended): "A Mentoring Community Project will be created and piloted this year and fully implemented in 2007."

Environment. As the Team's energy was fresh, this external scan proved meaty. A web search led to a referral resource at Mentors Peer Resources Guide online. Those web pages contained lists of colleges and universities with mentoring programs of many configurations[46]. Analysis of external scans determined that mentoring programs appeared most advanced at the following peer and aspiration institutions: Michigan, Virginia, Maryland, and Rutgers. Michigan State was more valuable as an information clearing house than model for actual mentoring programs – more talk than walk. Wisconsin enjoyed a reputation of being advanced in strategic planning and evaluation activities, and so their models were reported to be of high quality, and

---

[46] Many programs were identified by a memorable acronym, in accordance with sound marketing principles, i.e., The Women in Science and Engineering (WISE) at Illinois; Learning Integrated Freshman Experience (LIFE) at Delaware; NIA (meaning "purpose" in Kiswahili) at Buffalo; Knowledge, Excellence, Wisdom, & Learning (KEWL) at Penn State; and Mentoring & International eXchange (MIX) at Virginia. I think these were some of the coolest acronyms and nomenclatures unearthed in any of our scans.

perhaps customizable for Pitt's purposes. The closest parallel to current efforts was happening at Ohio State.

Inputs: *Student attributes / Assumptions*. As with the e-portfolios outcome, planners deferred to assumptions about the value of mentoring over an actual measurement of student wants and needs for mentorship programs. Management may have been correct and justified, albeit paternalistically, in seeking effective mentorship and networking programs that students might grow to value.

Assessment Team activity. As with e-portfolios, the Team began evaluation with a search for a comprehensive definition. The Team found that mentorship arrangements for students can include peer to peer, staff to student, faculty to student, and student to youth. Mentoring programs have evolved for numerous community interests.

The Team considered that Student Affairs might not have to create its own new systems, but rather can network with and customize what was already available. The Team pointed out that Student Affairs could create a *general* mentoring community web page with links of interest and value to Pitt students, as a way to supplement new formal mentoring communities at Pitt.

The Team's external scan located commentary that mentoring program benefits included 1) training knowledgeable leaders at every level of the organization; 2) increased ability to recruit talented employees; 3) facilitating "rapid gain" in organizational knowledge for new employees; 4) promoting higher retention of skilled employees who want to develop professionally and personally; 5) increasing satisfaction for mentors able to influence and participate in the professional and personal development of a less experienced person in the organization, 6) providing opportunity to learn from the mentee (Michigan State University, M.E.N.T.O.R.S., 2006). For student participants, mentorship could offer a means to enhance

current job performance and develop competencies essential for future career development, including self-confidence, managerial and leadership skills, insights about themselves, problem-solving skills, and organizational understanding (Michigan State University, M.E.N.T.O.R.S., 2006).

Outputs: *Programs*. In internal scans, the Team identified an issue that current mentoring programs represented single events, as opposed to the ongoing and process-based relationships to which a true mentoring community could aspire. The Team identified existing programs at Pitt as possible foundations for a viable mentoring community, in line with defined Student Affairs outcomes: Pathway to Professions; Pitt Career Network (> 1,100 alumni); College of Arts and Sciences freshman mentors; Greek academic mentoring.

The Team located campus precedent for mentoring programs at the International Office; the "Freedom" program, where seniors trained as "conductors" guided underclassmen through a workable schedule; and programs for under-represented groups, including a proposal for a mentoring program for commuter and other non-traditional students not residing in the residence halls. The Team thought that Pitt's LLC's presented a unique and successful spin on academic and discipline-based mentorships. The program seemed consistent with the Pitt student body's reputation of having a share of energetic scholars with interests in research. LLC's require dedicated mentors to serve as facilitators and role models.

Outcome (Actual). The assessment campaign informed new policy, as Student Affairs management moved forward with pilot mentoring programs, including a monthly round-table at the newest residence hall. This result was favorable for evaluators who might find validation in seeing recommendations transform into policy, manifesting purposeful collaboration between decision-makers and evaluators.

142

<u>Meta-evaluation</u>.

Figure 24   **Logic model depiction of evaluation of mentoring outcome**
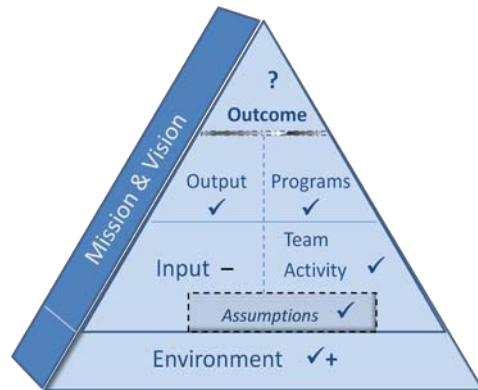
The Team did a thorough job investigating this outcome, except for being thin on the needs audit. The utility of mentorship programs was presumed – a correct assumption in my opinion. The Team conducted an exemplary external scan.

Achievement of the outcome remains in question. It appears that a pilot program has failed, but this does not mean that other such programs would not succeed with better preparation and publicity. Participation is the key element.

Student Affairs did not succeed in the intended full implementation by 2007, and should restate this outcome with a new target date – if and only if needs audits confirm student needs and desires for more formal and systematic mentorship programs than the organic ones that happen naturally. Evaluators and planners must also be ready to deliver surveys in the future that measure how young alumni believe they have benefitted from these programs.

B.2   INTERNSHIPS (#8)

Outcome (Intended): "Every Student Affairs unit will create non-paid internships to create additional learning experiences."

Environment. Case studies found that interns should always be compensated, primarily on grounds of "moral obligation to train students in their discipline" and that "agencies which do not pay interns would have to upgrade their training programs to still attract interns" (Geist, 1977, p. 40). In a longitudinal study, a team at Southwest Missouri State found conversely that interns considered "people skills" to be their most valuable gain from internships, and "money was not found to be a major factor in the respondents' involvement with the internship program" (Cook, Parker & Pettijohn, 2004). At SWMSU, like at Pitt, receptiveness toward unpaid internship programs was a function of organizational culture.

Inputs: *Student attributes*. Students' preference for paid internships over unpaid was presumed. However, budgetary constraints resulted in the next-best alternative: unpaid ones.

A better way to conduct a needs assessment on unpaid internships would be 1) to survey a selected group of prospective interns on their incentives and expectations and 2) to interview students who had already participated in such internships. This would enable teams and units to develop a student profile, marketing campaign, sponsorship / partnerships, training programs, and curricular credits or other incentives.

Assumptions.   The primary hurdle for unpaid internships was providing an opportunity and reward that could compete with paid internships and jobs on or off campus. Unpaid internships involve high opportunity cost. Academic credit was considered as one possible resolution. Cross-training throughout different Student Affairs units might provide exposure to a variety of related disciplines. Another possibility was future consideration for paid positions,

similar to an office apprenticeship. Any such resolution would require closer collaboration with partners inside and outside Student Affairs.

Assessment Team activity. In a report submitted to the Dean in early February 2006, the Team noted that its departmental audit revealed that most departments already offered such internships, although not all were unpaid. Student Affairs had already fulfilled this particular outcome and only needed to continue monitoring internships, and perhaps consider conducting follow-up correspondence to learn how these internships transitioned into careers and graduate study. A related recommendation was to find ways to overcome or reduce competition between unpaid internships and paid positions through different types of incentives, possibly contractual.

Outputs: *Programs.* The team conducted an internal scan of all seven units to catalog currently sponsored non-paid internships in Student Affairs. Among active programs identified by the Team, Student Life offered stipendiary and non-paid internships for a dozen organizations. Residence Life offered one internship, an assistant to the Associate Director. While Student Health had no official internships, a few programs accomplished similar aims. Because the information was sensitive, the University Student Judicial System did not offer any internships, as only paid contractual workers could be held accountable for confidential documents and processes. Disability Services offered two internship opportunities. The Office of International Services sponsored an intern through the International Research and Exchanges Board. Another intern volunteered with the OIS Programming Section for eight months and leveraged this experience towards employment first as a chaperone for a local high school experiential learning program in Italy and subsequently as a high school teacher in Italy. This student's example was precisely the type of success story that Student Affairs had envisioned with an internship outcome. In addition, OIS Ambassador Corps has maintained a corps of about

20 volunteers who help out with program planning, logistics and implementation. These programs dovetail with the "global learning experiences" outcome.

Career Services offered wide range of learning opportunities through three internships within the department, plus ten to twenty Emerging Leaders Peer Facilitators. Career Services occasionally accepted interns from graduate programs in counseling and higher education administration to assist with day-to-day functions. Counseling Center sponsored one unpaid Masters level intern from the University of Pittsburgh School of Social Work. Finally, internships were a feature of a new Office of Cross-Cultural and Civic Leadership (O3CL) established in August 2006, after issue of the Team's evaluation report on internships.

Outcome (Actual). The Team's efforts achieved confirmation of an active outcome. The next step would be sustaining internship programs, which seemed to be thriving.

Meta-evaluation.

Figure 25   Logic model depiction of evaluation of internship outcome
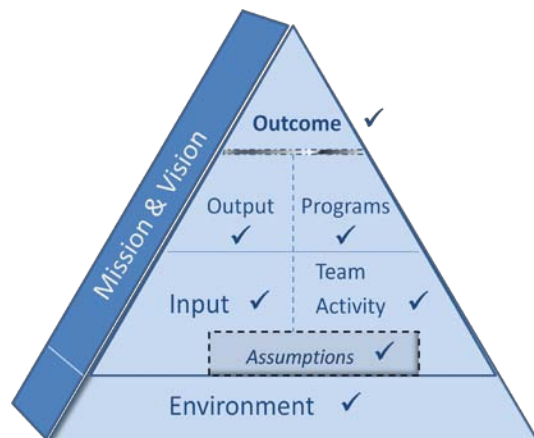
The Team did a thorough job investigating this outcome, except for again being thin on the needs audit. Locating solid participation in internship programs throughout the department and Team interviews with participants yielded a check mark in the "input" slot. The Team's external scan could have been broader, and centered more on peer and aspirant institutes for

direct comparison than generic testimony from literature reviews. I was tempted to add a "plus" to the check mark in the "program / activity" box, as the Team located active internship programs galore, with considerable participation. Some of these participants have already offered positive feedback on their experiences, validating this outcome further.

I considered internships the "feel good" outcome of the campaign – the Team's low-hanging peach.

B.3   PITT PROMISE (#5)

Outcome (Intended): "Students will be constantly reminded of the Pitt Promise to Civility."

Environment. We could have contacted military academies for descriptions of how their students appreciate recitation of codes and disciplinary exercises. Notre Dame might have been another good source of information for its renowned honor code. An overlap between "environment" and "assumptions" might confirm suspicions that Pitt's profile is too different from those campus profiles for meaningful direct comparison.

Inputs: The closest approach the Team made towards an internal scan was informal conversations between me and RD's who described the tone and mood in residence halls. They described morale as high and occasionally infectious. I shared with the Team my belief that campus students would be receptive to this outcome. Perhaps we could have done more than trust hearsay, but RA's and RD's are exactly the type of agents evaluators enlist to conduct an internal scan, because RA and RD familiarity and rapport with students could result in data easily and willingly.

Assumptions.   Team assumptions included belief in the inherent goodness of honor codes, not cheating, and communal goodwill – such as respecting privacy and "quiet hours". Gathering evidence on students' belief in citizenship could have seemed paternalistic and intrusive, especially with danger of students getting "survey fatigue". We thought the annual Quality of Life survey (in April) would cover citizenship enough for continuity with Team activity.

Assessment Team activity. The team could have taken formal and documented action to draw upon input variables. Academic models focusing on discipline and ritual exist on and outside the Pitt campus, and some of these principles are translatable if not universal. We could have interviewed students engaged in associations, such as the Black Action committee, Rainbow Alliance, and fraternities. Possibly a random sample of alumni could reveal how many could sing their alma mater, and whether its words still ring true and meaningfully. If we had conducted a thorough literature review, we could have cited student development theory on the value of promises, academic vows. Instead, a lot was left to assumption that a promise is a valuable outcome[47].

In a report submitted to the Dean in early February 2006, the Team reported finding that every department had some involvement with the Pitt Promise. If this appraisal seemed qualitative and fuzzy, we considered how to design quantitative or pattern-based measures of success in the future – for example by surveying recent graduates, conducting online focus groups, and cohort studies. Directors and planners may consider that type of follow-up in a subsequent evaluation phase.

---

[47] I made it a point to append a copy of the Pitt Promise with every agenda delivered electronically or on paper to the Assessment and Measurement Team in preparation for meetings.

Outputs. The Team's internal scan identified active programs on Student Affairs and unit levels, usually involving signage[48].

Outcome (Actual).   Like Pitt Pathway, the success of the Promise outcome may unfold over time. Evaluators may discover that years from now alumni remember the Pitt Promise by heart, and continue to derive meaning the way they might sing their alma mater after a sporting event, at reunion, or in other shared nostalgic moments.

Meta-evaluation.

Figure 26   Logic model depiction of evaluation of Pitt Promise outcome
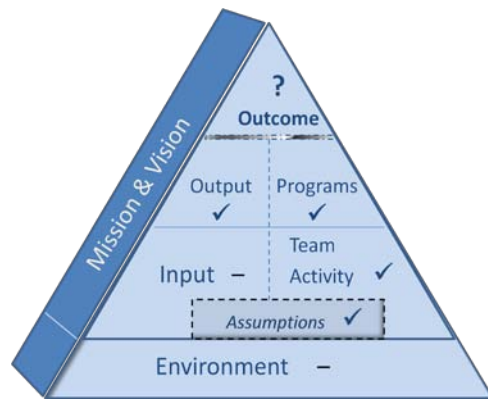
A lapse was neglecting to gather from literature and comparison institutes substantial evidence of the effect of promises and codes upon campus morale and alumni behavior. A key element of evaluation would be to measure participation, offering modest quantitative support on how students subscribe to this idea. Still, this has been a worthy outcome, and its process was well executed. For longer term impact, a possible ramification could be richer alumni loyalty and boosts in consequent donations.

---

[48] Another assertive example was that one RD required any advisee to recite the Pitt Promise before gaining admittance to her office.

The Pitt Promise: A Commitment to Civility

The University of Pittsburgh is committed to the advancement of learning and service to society. This is best accomplished in an atmosphere of mutual respect and civility, self-restraint, concern for others, and academic integrity. By choosing to join this community, I accept the obligation to live by these common values and commit myself to the following principles:

As a Pitt Student:
- *I will embrace the concept of a civil community which abhors violence, theft and exploitation of others,*
- *I will commit myself to the pursuit of knowledge with personal integrity and academic honesty,*
- *I will respect the sanctity of the learning environment and avoid disruptive and deceitful behavior toward other members of the campus community,*
- *I will support a culture of diversity by respecting the rights of those who differ from myself,*
- *I will contribute to the development of a caring community where compassion for others and freedom of thought and expression are valued,*
- *I will honor, challenge and contribute to the scholarly heritage left by those who preceded me and work to leave this a better place for those who follow.*

By endorsing these common principles, I accept a moral obligation to behave in ways that contribute to a civil campus environment and resolve to support this behavior in others. This commitment to civility is my promise to the University of Pittsburgh and its community of scholars.

B.4   LISTEN TO STUDENTS (#3)

Outcome (Intended): "Students will feel that we have listened to them."

Assumptions. The Team accepted "common knowledge" that interviews and focus groups could derive a deeper and more interactive human element from generic data, by techniques comparable to an immediate, journalistic form of participant-observation. These methods seemed a compelling way to achieve the "listening" outcome. Patton (1990), Weiss (1997) and other qualitative research experts corroborate our assumptions and actions in

retrospect; these same authors offer salient points to guide assumptions for future evaluation phases.

Inputs: *Student attributes.* Exploring this outcome represented an instance where the Team actively gathered input from students, thereby conducting a needs audit often missing from outcomes explored previously.

Assessment Team activity. The Team recognized interviews and focus groups as a primary vehicle to achieve this "listening" outcome, to gather opinions that might develop into needs audits, and to drive activities that would measure and fulfill other outcomes in tandem. For those reasons, the Team devoted time, energy, and research into strategizing purposeful interview and focus group sessions.

One task was to draft interview scripts for fresh sessions. Pilot tests among Resident Directors (RD's) yielded revisions to an earlier draft of interview questions and indicated that the finalized instrument would likely invite participation with minimal time commitment (of less than 10 minutes) per respondent. The Team proceeded with interviews. Six sub-teams conducted initial interviews with approximately 20 participants each from various locales on campus to capture a broad spectrum of the student body. The Team agreed that the described technique would effectively draw a random sample that fairly skimmed the surface of collective opinion among the student residential population. Interviewers, usually in pairs, approached random students in key locations, including the Student Health Pharmacy and the William Pitt Union basement (where the not-yet-refurbished game room was located). Some pairs manned tables that students approached voluntarily. International Services conducted interviews in their office waiting room immediately after the students had received services. Although that technique compromised the randomness of the sample, OIS gathered useful, representative data.

Usually interviewers read from a standard printed script and recorded students' responses. Other times the paper instrument was given to students, who filled in handwritten responses.

Team members reported being able to achieve modest demographic diversity, although the only demographic details tracked were gender and undergraduate / graduate status. Out of a total sample of 106 students, we drew 81 undergraduate responses – 32 female, 39 male and 10 not defined – plus an additional 25 responses from graduate students, for whom gender was not tracked. Although the undergraduate figure represented about 1.5% of the on-campus population, we had no reason to doubt this sample's reliable representation[49] of campus sentiment. Separations by gender and graduate status would have enabled follow-up evaluations to explore issues particular to certain groups, had we observed significant disparity.

Findings revealed a fairly uniform response among all groups. Qualitative descriptions were seen as a way to provide precedent for future research, including focus groups over the subsequent few weeks. Comments fell into patterns and categories, listed in order of frequency in Tables 8 and 9 in the outputs sub-section of this sub-chapter. The summary report of ~ 120 interviews was completed and submitted to the Dean of Student Affairs in mid-March. An Executive Summary of Student Interviews on Baseline Outcomes, released March 22, 2006 to the Dean summarized findings from the sequence of interviews conducted in February and March 2006 by members of the Student Affairs Assessment Team.

From interview findings, the Team devised a draft script and list of questions for upcoming focus groups.

---

[49] Online calculators (Creative Research Systems, 2009, http://www.surveysystem.com/sscalc.htm) indicate that the campus population requires a sample of about 350 to achieve 95% confidence levels for 5% confidence intervals. This sample achieves between a 9.5% and 10.5% interval with 95% confidence.

Outputs: The following tables show findings from student interviews conducted in February and March 2006. (Executive Summary of Student Interviews on Baseline Outcomes, March 22, 2006, p. 4).

Table 8   Students' suggestions to improve their Pitt experience

Issue / category of comment	Frequency (N= 81)
Better food (less fried) / cheaper / vegetarian	18.5%
Inconsistent enforcement of rules	11%
Advertise programs more	10%
Academic standards / advisor quality inconsistent	10%
Department coordination & communication	7.5%
Transportation & shuttles	7.5%

Table 9   Students' suggestions to improve Student Affairs

Issue / category	Frequency (N= 81)
Advertise, publicize more (+ Pitt News)	11%
More activities / interaction / events	11%
Integration & bureaucracy: waiting, scheduling, attention, care	7.5%
Student input	7.5%

The Team planned to follow up sessions and reports with additional surveys and/or focus groups (also conducted by student representatives from Emerging Leaders using protocols designed by the Team). Fall 2006 was considered as a target date, and recommendations blossomed into a division-wide effort with each department defining expectations. Thus, "listening" initiatives immediately resulted in new and reconfigured programs.

The Team concluded that "this exercise has been effective in continuing to give students a voice in suggesting and influencing policy… However, only the implementation of progressive

153

policy and programming that is sensitive to their needs will offer them proof that "we have listened to them" (Executive Summary of Student Interviews on Baseline Outcomes, March 22, 2006, p. 7).

Programs. All units collected surveys on office visits and manned hotlines for questions and feedback. Several had "suggestion boxes". A reliable system of RD's and RA's guaranteed a listening ear to student concerns.

Outcome (Actual). We listened, and the Team drew evidence that students appreciated the Student Affairs department's efforts. There emerged a new collaborative program to conduct regular interviews and focus groups. A research limitation was that our evaluative approach enlisted feedback from students who were bustling about outside their rooms and presumably active on campus. Interviews have little access to shy and private students, who may have serious concerns that Student Affairs could address if voiced.

Meta-evaluation.

Figure 27   **Logic model depiction of evaluation of listening outcome**
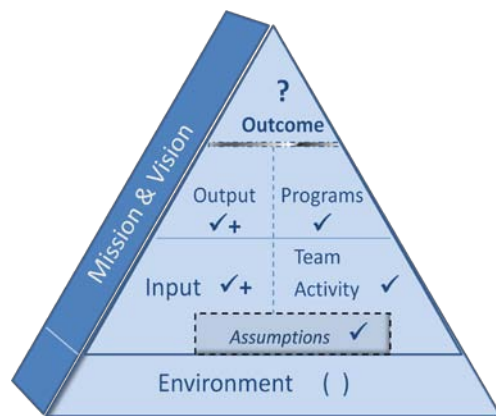
Evaluating this outcome provided residual evidence towards measurement of two other outcomes, #1 on excellence and #2 on Student Affairs performance. This notion of intertwined outcomes and processes illustrates the logic of Paulston's flexible mapping, rather than rigid characterizations of processes and prescribed measurements. Linkages to other measurable

154

outcomes reveal that "listening to students" accomplished even more than an outcome in itself. A significant finding is that if the process of listening to students is inherently good, their feedback can and should also inform fulfillment of other outcomes.

I would defend our Assessment Team's stance in that follow-ups among random respondents were the best we could do without intrusion, and we probably succeeded in gathering the best available data. A whole representation is realistic only in an ideal world of unlimited collection time and guileless interaction, but particularly unrealistic with populations of student wallflowers who prefer not to be bothered. Perhaps a realistic compromise is leaving invitations open for students to volunteer e-mails or calls – although this will result in what I have called the "usual suspect" response.

The intended outcome was not only for us to listen, but from a student standpoint to know that their voices made a difference in formulating programs. I think that increased participation and open communication between students and Student Affairs units will measure achievement of this outcome over time.

I reiterate that in the shorter term, results of this process have already become valuable inputs for other outcomes. If the Team indeed covered all bases and gathered valuable information, a boon to investigating this outcome was its segue into other outcomes. The logic model would show a flow of information from the "outcome" apex not only down-slope to inform evaluators in future phases how to listen to students attentively, but also across to other clusters of pyramid models – feeding into their "assumption" and "activity" blocks. The figure below clarifies that flow of knowledge, represented by green arrows.

155

Figure 28   Flow of knowledge from achieved outcome to future phases

and to other intended outcomes
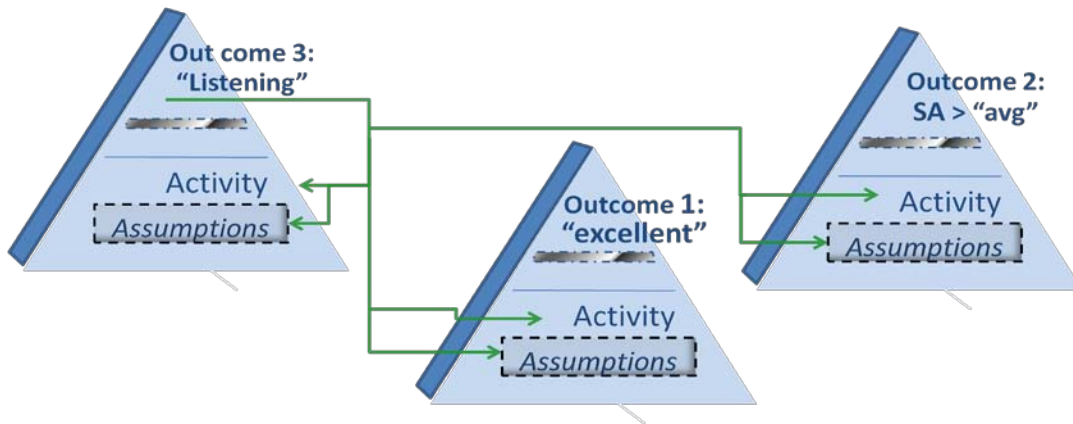
B.5   EXCELLENT EXPERIENCE (#1)

Outcome (Intended): "Students will rate their Pitt experience excellent."

Inputs: *Student attributes*. The Team used a valuable internal source of data to launch investigation of this outcome: results from the annual Quality of Life survey administered by Residence Life. I would argue that Student Affairs' most plentiful resource on student inputs, needs audits, and collective profiles was and is that survey.

Assumptions. The primary assumption was that survey respondents fairly represented the student body. Quantitative analysis, sample sizes, percentages, and measures of central tendency indicated the validity of data and of this assumption.

Assessment Team activity. This outcome illustrates two interesting sub-topics: 1) An assessment team should take advantage of research done concurrently in the same department, as this conserves efficiency; and 2) research may involve two or more interrelated outcomes.

Fulfilling outcome #3, listening to students, resulted in much data on how students valued their college experience.

Outputs: *Programs.* The Team located scant programming geared directly at measuring excellence or general quality. One datum was diminished attendance at an annual Fall Orientation.

Participation / Results. The table below offers a quantitative rating of a sample of Pitt campus residents' appraisal of their college experience (ranging from 1 to 5).

**Table 10   Students' assessment of their Pitt experience**

Average	Median	Mode
4.21	4	4

(n = 106; source: Executive Summary of Student Interviews on Baseline Outcomes, March 22, 2006, p. 3)

Outcome (Actual). The Team answered questions raised by this outcome with a single snapshot from results of the Quality of Life survey. The most important conclusion from this research was establishing a baseline, at 4.21. While not excellent, this score translated to "very good". The key from now on would be to re-evaluate this collective score every year scientifically, using consistent assumptions and calculation methods.

<u>Meta-evaluation</u>.

Figure 29   **Logic model depiction of evaluation of excellence outcome**
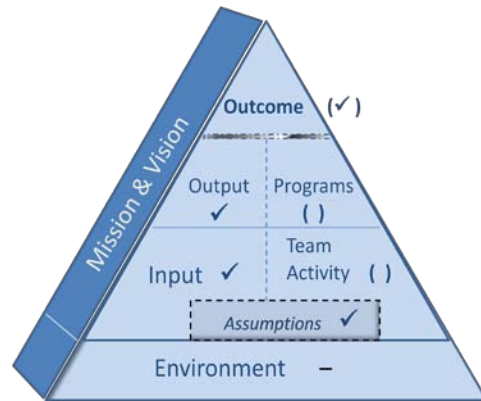
The Team performed this task efficiently and methodically. However, an omission was comparisons to other IHE's of similar profiles. To aspire towards excellence, it follows that the Team should have done a scan of aspiration institutes. The fact that Student Affairs has defined a set of "aspiration institutes" implies that excellence has not yet been achieved at Pitt, while academic rivals have reached a level closer to excellence. Thorough evaluation of this outcome could go deeper than hearing most students say, "Yes, I like it here." For future phases the Team must do more to validate this perception. While finding a benchmark point was a valuable accomplishment, it is a self-contained and incomplete measure without linkage to peer and aspiration institutes' visions of excellence. After all, statement of mission refers to "best collegiate experience", but the Team did not establish a standard to which "best" compares.

The logic model reveals another missing piece in how students' perceptions relate to programming. Analysis of this outcome did not establish causality whether campus programming added value to student perceptions and contentment. Good news is that the Team could request access to quantitative findings gathered by Residence Life and other units for detailed benchmarking. A way to test causality further is by comparing impressions of campus programming among non-participants to a control group as represented by Quality of Life survey

respondents. A null hypothesis could be that programming wallflowers or abstentions have loved their Pitt experience.

B.6   STUDENT AFFAIRS RATED ABOVE AVERAGE (#2)

Outcome (Intended): "Students will rate all Division of Student Affairs programs and services above average."

Environment.   As in the "excellence" outcome, the Team neglected to draw knowledge from peer and aspiration institutes on this outcome. That comparison may be pertinent depending on interpretations of "above average" as a raw quantitative score, as a baseline (my favored interpretation), or as comparison with peer and aspirant institutes' available evaluations of their Student Affairs departments.

Inputs: A facet not emphasized strongly enough in Team investigations was diversity. If the Quality of Life survey asked more details about demographics, hobbies, etc., then planners could have observed correlations between Student Affairs performance and specialized needs. One problem was that some lines of questioning, such as financial status and medical / handicap issues, could be intrusive or perhaps even illegal for units to track. The bottom line is that different groups and sub-groups could have rated various services differently, but activities did not gather variable distinctions for comparative analysis.

Assumptions.   A risky assumption related to data limitations described above was that feedback represented consensus. Displeased students might remain silent and unmeasured. A related assumption is that RA's would encourage students to speak out instead of simmer, and that on a local and organic level, RA's had been trained on how to appease unhappy residents

159

and minimize conflict. Leaving out those students could affect validity of samples used to determine Student Affairs performance ratings and averages.

Assessment Team activity. A first step that arose in an early (November 2005) meeting was to define "what is Student Affairs" and to achieve consensus. Internal scans yielded findings upon inquiry whether instruments (particularly a suggestion box) were in place to collect complaints and compliments that would measure Student Affairs services. The report submitted on May 18, 2006 to the Dean of Student Affairs covered three outcomes together, under the heading of "performance outcomes". To establish baselines, that report quantified (by ratings based on the familiar 1 to 5 scale) findings from student interviews conducted in February and March 2006, as shown in the following table:

Table 11   Average ratings by random students of Student Affairs services

Student Affairs unit	Average (Mean)	Median	Mode	# of respondents
Career Services	3.84	4	4	64
Counseling Center	3.83	4	4	36
Disability Resources	3.44	4	4	9
International Services	4.06	4	4	33
Residence Life	3.71	4	4	83
Student Health	3.95	4	5	80
Judicial System	2.83	3	4	12
Student Life	4.12	4	4	81
Overall	**3.87**	4	4	106

(source: Executive Summary of Student Interviews on Baseline Outcomes, March 22, 2006, p. 3)

Total (n=106) represented instances, rather than a unique number of respondents, since some students had experience with more than one unit.

160

Outputs: The Team's investigations identified methods and actions, some of which may qualify as "programs", to measure unit-level and overall Student Affairs performance.

Outcome (Actual).   The Team audit revealed that Pitt Student Affairs units regularly engaged in measures related to staff performance. The May 2006 report served as a compiled baseline, upon which each unit could be encouraged to seek improvement both independently and collaboratively within Student Affairs. Process would likely center around communication, including a sharing of effective practices and about policy that had not resulted in intended goals.

For Student Affairs services, a close cluster around an overall average score of 3.87 equated to a rating leaning more towards "very good" than "good / adequate". This overall score indicated that students were usually pleased, but there were areas of improvement. Statistics indicated that this sample of students considered department and unit services to be above average, and this may have reflected overall campus sentiment.   Only Student Life and International Services achieved a rating better than "above average". The Team offered kudos to those two units, and noted impetus to keep up the good work, while other Student Affairs units should conduct formative steps to boost students' impressions of services and programs. The Team diagnosed such next steps as monitoring progress from newly established baselines. Data gathering methods could include a combination of focus groups, interviews, surveys, and directorial evaluation of outcomes. Success would likely require collaboration between independent, unit level strategies and Team oversight of standard models and customized solutions.

Meta-evaluation.

Figure 30   Logic model depiction of evaluation of Student Affairs performance outcome
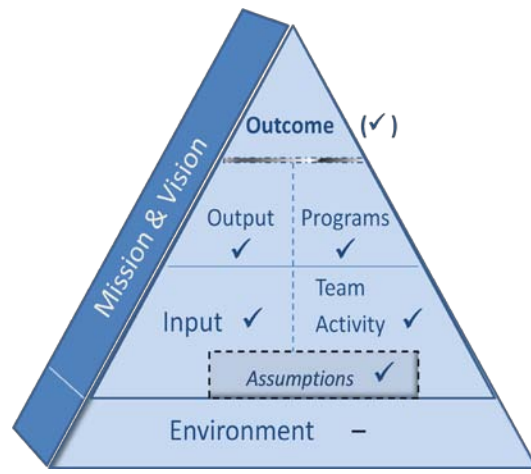
The Team evaluated this outcome efficiently and methodically by drawing from existing data sources, while again lacking comparisons to other IHE's. Internal scans were comprehensive in auditing current programming. Exploding the "input" block of the logic model might encourage evaluators to seek feedback from students who had been non-participants, perhaps by one-on-one sessions conducted by RA "agents". Team assumptions about this outcome seemed clear, but it remains debatable that quantitative scores above "3" denote achievement of this outcome. Planners could consider whether to determine incremental improvements that aspire for "excellence", now that "above average" has nominally been achieved across the boards.

I think that this outcome, of all 14, best captures the essence and minimum goal of holistic assessment. If a "culture of assessment" accomplishes no other goals, it must encourage units to hold themselves accountable for programming performance and quality. The Team did an admirable job establishing quantitative baselines for each unit. The structure and compartmentalization offered by this study's logic model could help evaluators and planners

accomplish those tasks on levels that now move from departmental and Team accountability to unit accountability.

B.7   CUT PROGRAMS (#6)

Outcome (Intended): "All programs that do not adequately meet the needs of student will be modified or eradicated."

Assumptions.   The underlying assumption for this outcome, beyond theoretical need for periodic summative review, was an efficiency argument: how best to serve students with limited budgets and competing resources.

Input. *Participation.*   The Team trusted units to have determined adequate participation by students in programs, justifying costs. Input details and attributes did not come into play other than an understanding that if students liked a program, it would continue.

Assessment Team activity. The Team found this outcome to be the most contentious of Student Affairs performance outcomes investigated in tandem and compiled into a report submitted on May 18, 2006.   Assessment Team members inventoried their respective departments and requested a brief status report from each Director. This was essentially a prompting for lists of summative judgments.

Outputs: Directors' replies often were terse. Investigating this outcome was not intended to be intrusive, but Directors perceived it that way and evidently were not pleased. Despite resistance, the Team culled information on unit-level programs and their fates.

Outcome (Actual). Our audit revealed that some Student Affairs units were actively engaged in achieving outcomes related to staff performance and concentrated funding and

163

staffing (including volunteers) on programs that measurably enhanced students' experiences on campus. The ensuing report served as a modest compiled baseline, upon which each unit could seek improvements both independently and collaboratively within Student Affairs. As with other performance outcomes, collaboration would likely center around communication, such that a unit would not find itself referring students to programs that no longer existed or would be cut soon – an awkward circumstance.

However, outcome #6 on summative evaluation of programs provoked some dissention throughout the department, in that Directors felt autonomous in defending certain programs and policies and in deciding how other programs might need restructuring. This outcome trod delicate lines between open communication and diplomacy. Perhaps this assertive outcome ushered in a "culture of [joint] accountability" more quickly than a gentler approach, which would have allowed independent-minded directors to hide behind procedure. This outcome broke down walls of enclosure and secrecy. A likely reason was suspicion about budgetary impact upon future allocations. Did open communication and mutual trust represent a worthy trade for less autonomy?

The next steps should involve monitoring progress from newly established baselines. Techniques should include a combination of focus groups, interviews, surveys, and directorial evaluation of outcomes against inputs ranging from 1) collaboration to independent unit level strategies and 2) standard models to customized solutions.

Meta-evaluation.   Early on, the logic model would have erupted into something like the figure below.

Figure 31   Logic model depiction of summative evaluation outcome eruption
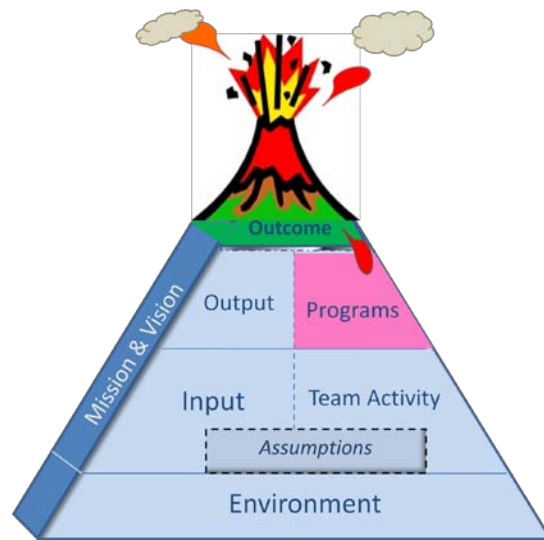
Impact. This directive was initially a political firestorm. Directors resented sharing accountability as a challenge to autonomy and evidence of micro-management. Results of this evaluation demonstrate how, occasionally, feelings get hurt. Someone could regret that a program got scrapped for collective unpopularity or practical infeasibility. In other cases, an unscrupulous leader could point to the evaluator and announce, "His data told us to kill the program!"

However, I believe that this outcome has gained credence as the Team transitioned into an advisory role. By organizational definition, Directors could now enlist the Team's expertise by independent discretion and timing, according to their own reckoning of unit budgets, and instead of a "forced" campaign across student affairs to boost efficiency and allocate resources. The logic model might be adapted to resemble the figure below.

Figure 32   A reconfiguration of summative evaluation outcome logic model (Team as advisory)
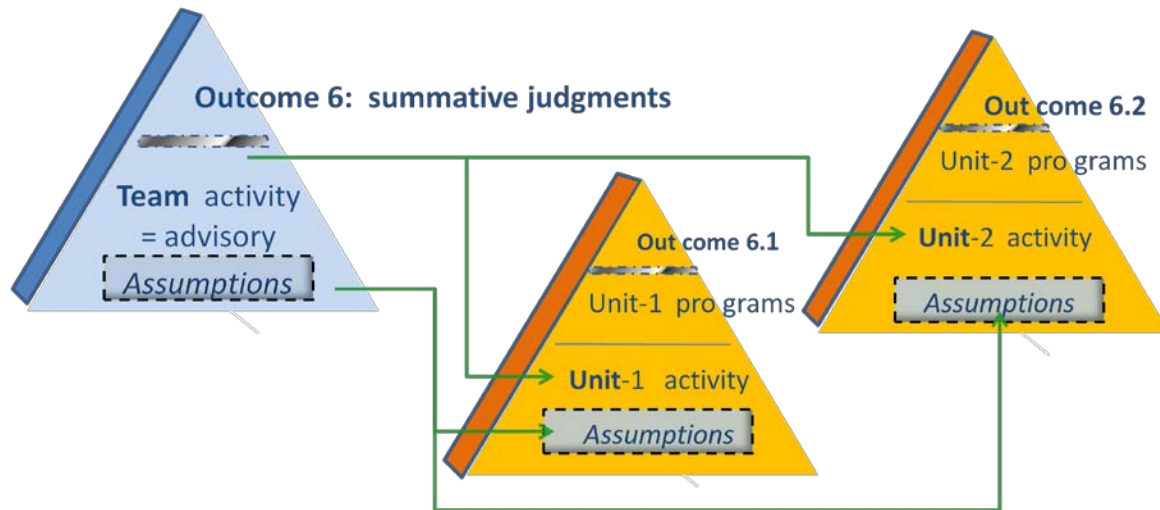
Ultimately this outcome granted authority to the Team for collaboration, while preserving autonomy to units and their Directors. That notion has been essential for forging future policy from current evaluations.

B.8   PARENTS CONDUIT (#14)

Outcome (Intended): "A conduit of information for parents will be created and utilized."

Environment. The Team's external scans focused mainly on peer comparisons. Explorations revealed many informative websites with FAQ's, local events, and even a bilingual portal at Michigan. Consensus, as reflected in entertaining phrasing from Assessment Team Meeting Minutes, February 7, 2006, was that "we're not good at parent" (p. 2) as some of our peers.

Inputs: *Student attributes*. Student Affairs professionals frequently complained about "helicopter parents". Closer relationships between students and their parents than was observed among prior generations of college students may have been part of the impetus for this outcome.

Assumptions. The primary assumption was parents' comfort with and interest in technological solutions that might replace personal and direct contact. Would the promptness of self-service for answers and information be a worthy compromise for parents perhaps expecting personal contact? Proactive measures such as newsletters and e-mail messages might reduce Student Affairs phone call traffic and limit contact to actual emergencies and exceptions.

Assessment Team activity. The Team conducted internal and external scans on programs used to keep parents informed. The internal audit consisted of three questions to each Student Affairs unit:

1. What method(s) if any, does the unit utilize to contact parents?
2. Who typically in the unit makes contact?
3. How often does your unit contact parents?

Co-chairs compiled findings into a report submitted on February 21, 2006. A revised report was issued on May 19 upon results of a second external scan. Based on audits of active programming at Pitt and comparison colleges, the Team made recommendations on strategy. One suggestion was tapping into the *Pitt News* as a good outlet for Student Affairs information, since many parents had subscriptions to the paper. Another idea was to adapt a periodic newsletter into a parents' edition. A "Parents' Page" on the Student Affairs website might follow a three-pronged approach to inform parents: 1) Recruitment: "Send your kid here!" 2) Quality of life: "You made the right decision!" Supplemental information and statistics might address safety, accommodations, and rankings among other qualities, such as Pittsburgh as a legitimate cosmopolitan and cultural center. 3) Problem solving and commitment to an ongoing relationship: "We care!"

Outputs: *Programs*. On unit and divisional levels, the Team identified numerous programs and actions in support of this outcome.

Outcome (Actual). Student Affairs offered a commendable array of programs, services and activities intended to keep parents informed and to include them in students' progress. Audits revealed that Pitt Student Affairs units contacted parents usually on the basis of particular events or incidents, such as a violation or illness. Current contact with parents could be described either as "reactive" to circumstances or "ordinary" in terms of scheduled events. Not coincidentally, the Team identified departmental "strong points" in keeping parents informed about events like Freshman Family Weekend and orientations.

Among proactive approaches to keeping parents informed, a newsletter to Pitt parents seemed like an unobtrusive, affordable, informative solution. The Marketing & Communications Team's launch of a parents' version of a monthly newsletter turned out to be one of the most visible and applauded outputs of evaluation and related processes. An online survey, whose link was listed in the newsletter, collected feedback from readers. Survey response had been infrequent, and so the Assessment and Marketing Teams expected to collaborate on ways to encourage participation or to design and publicize other channels to gather feedback.

While good programs were in place for episodic schedules, the Team advised Student Affairs leaders to look to models presented by peer institutes on creating efficient ways of steady communication with parents. The Team speculated that parents may appreciate the gesture of a continual flow of information, rather than incident-based correspondence. Online functionality would allow curious parents to browse as they please, and plan accordingly – rather than relying on Student Affairs services to keep them informed through "bulk" approaches. Information available via (e-)newsletter or website could give parents notice *before* an incident might escalate into "crisis" status. If policy is posted, then parents would less likely be surprised by policy enforcement, especially in the case of violations.

Maryland and Michigan offered the most practical comparisons among the ones observed, while Ball State featured a comprehensive and interactive model that might offer more than is actually needed. The Team did not investigate why some peers, particularly Rutgers and Penn State had not yet made concerted efforts on parent information channels, at minimum with a comprehensive website. The question arose how such efforts may or may not comprise part of those peer institutes' Student Affairs strategies.

Meta-evaluation.

Figure 33   Logic model depiction of evaluation of parents conduit outcome
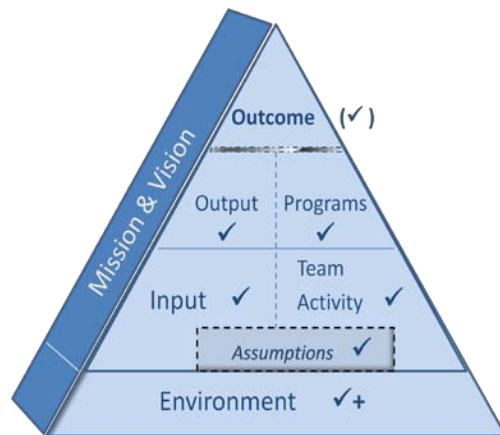
This outcome represents a comprehensive and effective evaluation by the Team. The environmental scan revealed valuable information on current programming and rich comparisons with peer models. The internal scan recognized a peculiar and (until the late 1990's) unprecedented character of "millennial generation" students, whose closer relationships with parents were a critical variable in designing useful programs.

Without a logic model, the Team somehow followed a logical progression from layer to layer of process. Using a model for successive phases could guide specific steps, fill in gaps, allow sound conclusions to be preserved and developed further, and help evaluators and planners know when it is time to gather data that might balance various sentiments.

169

B.9   FACULTY TOOLS (#13)

Outcome (Intended): "Faculty will be provided with additional tools to better understand and assist students."

Environment.   The Team located "powerful approaches" to faculty liaisons at four aspiration institutes and one peer, programs worth emulating. The Team considered a worthy topic of further study to be the role of faculty advisors in the functions and success of Living Learning Programs and Communities (LLC's), especially as Pitt planned greater ongoing investment in these organizations. The Team was unable to speculate how relations between Academic Affairs and Student Affairs at other campuses might echo tense relations and occasional disharmony at Pitt. One of Student Affairs' goals in creating a "culture of assessment" had been to transform skepticism and presumptions about Student Affairs by "traditional" academic units into mutually beneficial collaboration.

Inputs: *Student attributes.* The Quality of Life survey included questions on faculty interaction, by quantifiable and task-based measures such as frequency of offices visits and number of collaborations on projects, rather than depth of relationships. The Team conducted evaluation of this outcome from the standpoint of students and benefits they would derive from contact with faculty. Although the Team recognized faculty as fellow stakeholders, viewing faculty as another "input" could have resulted in well-rounded analysis.

Assumptions.   The Marketing Team shared commitment to promote the branding of Student Affairs, and to make our offices and functions recognizable at a glance. At the same time, teams agreed to be mindful to preserve distinctions from Academic Affairs. The Assessment Team realized the challenge of enabling faculty to be aware of those same distinctions between "baseline outcomes" and "learning outcomes". Of the two, how could

faculty priority not be learning outcomes? How could we convince faculty to *care* about baseline outcomes? An underlying assumption was that our prerequisite might be to convince faculty to care about Student Affairs at all, and then about our outcomes.

Assessment Team activity. As one of the few outcomes for which a complete evaluation report had not yet been issued, the "faculty" outcome became a top priority for our renewed assessment campaign in Fall 2006. So, this outcome spanned phases. The Team started the renewed campaign by reviewing notes from the February 7, 2006 minutes. Comments pertaining to this outcome focused on recent release of the new Student Affairs newsletter to faculty and other stakeholders. Participants expressed some concern that few responses had come to an on-line survey, until we deliberated about realistic expectations for such feedback. We realized that expectations could have been unrealistic – an important check and balance in evaluation.

The Team proceeded with internal and external scans of current efforts to inform faculty of Student Affairs activity and policy and to elicit feedback. These scans were modeled upon the prior spring term's productive scans of parental conduits, a fellow "stakeholder" outcome. A report submitted to the Dean of Student Affairs on February 6, 2007, explained the Team's findings and recommendations for this outcome. Scans found basis in that "existing research supports the widespread belief that student-faculty interactions are important to a student's collegiate experience" (Garrett & Zabriskie, 2004, p. 39).

Outputs: *Programs*. A report on December 4, 2006, covered a two-pronged internal scan: 1) current status of information channels between Student Affairs and faculty and 2) interaction between faculty and Pitt students. Results gave evidence of active programming intended to deliver "additional tools" for faculty. Student Life reported an impressive, lengthy

171

list of "viable methods and directives to integrate student and academic services effectively" ("Evaluation of faculty tools & communication outcomes", February 6, 2006, p. 1).

Outcome (Actual). Among new developments, the creation of O3CL and the newsletter represented positive steps towards the outcome of better faculty tools. Audits revealed that Pitt Student Affairs units regularly contacted and informed faculty on the basis either of 1) semester schedules and events – such as orientation and periodic themes – or 2) ongoing mentorship, advisory capacities, student support, and relationships. A more accurate statement of this outcome, in terms of an admirable status quo of faculty utilities, might be to improve and customize existing information channels and partnerships.

Still, Pitt Student Affairs efforts seemed to trail advancements by peers and aspiration institutes. The Team opined that UNC offered the most comprehensive and interactive faculty conduit models among the ones we observed. An advanced network like theirs may represent a long-term goal for Student Affairs at Pitt. Virginia may meanwhile provide a realistic short-term model for immediate implementation of improved faculty liaisons.

Meta-evaluation.

Figure 34   **Logic model depiction of evaluation of faculty tools outcome**
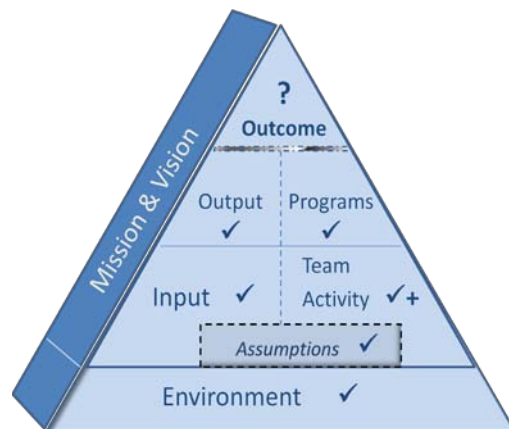
This was one of the thorniest outcomes, because faculty opinion of Student Affairs has ranged from advocacy to unmitigated scorn – even in the same department. The Team's approach was minimalist, such that enthusiastic faculty could participate while outreach would not inconvenience grouches (who could ignore or trash a mailing from Student Affairs).

Student Affairs could continue a minimalist campaign to achieve this outcome, and justify that providing tools satisfies mission and vision. This begs the question whether planners can do anything to make faculty care about student affairs. To address that issue, Student Affairs evaluators could consider adding a second layer to the logic model "input" block, to enlist faculty opinion and to conduct a needs audit of faculty. The fact that a Faculty Senate pilot program is on the table gives summative evidence that faculty might prove more receptive to active input from Student Affairs than this evaluation revealed – yet another example how and why evaluation does not terminate, but cycles back to successive phases..

B.10   ALCOHOL VIOLATIONS (#9)

Outcome (Intended): "The number of alcohol incidences will decrease."

Environment. Surely Pitt is not alone in this outcome. The Team should have explored literature and peer comparisons. The compliance component (i.e., drinking age, PLCB distribution laws, etc.) was accepted as a given – including enforcement by Pitt and city police (entities subject to separate and independent evaluations by their own discretion).

Inputs: *Student attributes.* Inputs and characteristics were self-defined: students cited with alcohol issues. It is debatable whether an evaluator should include student input and feedback in disciplinary outcomes, except in regard to fairness and due process. Analysis of

Outcome #3 on listening revealed that students generally thought USJS treated them fairly. The "input" block of this logic model is relatively simple.

Assumptions.   Originally the Team had considered that outcomes related to alcohol incidences and conduct violations might best be confined to Residence Life and Judicial Services as unit-level evaluations. In the process of evaluation, the Team heeded baselines and benchmarking from USJS and Residence Life archives, and we found that contexts and tracking frequently changed. Other issues included 1) on- versus off-campus violations and 2) charges versus sanctions. For those reasons, the Team considered a broad formative evaluation that might guide standardized and consistent policy.

Assessment Team activity.   Opinion arose that alcohol services shared among units needed more coordination with Judicial Services. The utility in place was MSB (mystudentbody.com), a subscription service that offered data reporting functions, longitudinal data, "utilization data", and a "resource binder". MSB was later one casualty of summative judgment and budgeting.

In this study's scope, the Team had not issued an evaluation report on this outcome. Still, Team research laid some useful groundwork and located programs intended to serve this outcome.

Outputs: *Programs.* Besides periodic awareness events, the primary program was PEAR (Personal Education, Assistance, & Referral) classes on alcohol use. PEAR delivers sanctions – possibly useful deterrent but a stretch in the definition of programming.

Pitt established another baseline for future benchmarking from a different angle: with its April 2007 participation, among 300 institutions, in an annual National College Health Assessment, one of whose topics included alcohol abuse. Results came in after my departure and

hence are technically out of scope for this study. Still, the results are worth sharing as an illustration of one complete cycle of evaluation – and which I recommend a logic model can help continue. The study found that Pitt students scored slightly higher than the national average of 38% on binge drinking, at 42.5% (Hart, 2008, p. 1). Follow up interviews by Pitt Student Health found that students ranked faculty third, ahead of parents and behind only medical staff and health educators, as potential advisors and confidantes on health-related issues including alcohol usage. These findings provoked collaboration between PLCB sponsors and Pitt faculty on workshops focusing on the effects of alcohol use and abuse. These findings also lend credence that the "faculty tools" outcome is indeed a very worthy one.

Participation. A critical benchmark would have been numbers of violations per student and instance, possibly with residence hall breakdowns. As USJS assimilated into Student Life, these requests lost urgency, or perhaps the analyst in charge of this program lacked confidence in available numbers.

<u>Outcome</u> (Actual). While much of this process happened outside Team activity but partially within Student Affairs, the process demonstrates how planners reacted to evidence and attempted positive change through evaluation and assessment that the Team had helped conduct and consolidate. I would recommend that the Team prepare to fit whatever data is available, including the annual National College Health Assessment survey results, into the "environment" piece of a logic model in evaluating programs related to alcohol use.

<u>Meta-evaluation</u>.

Figure 35   **Logic model depiction of evaluation of alcohol violations outcome**
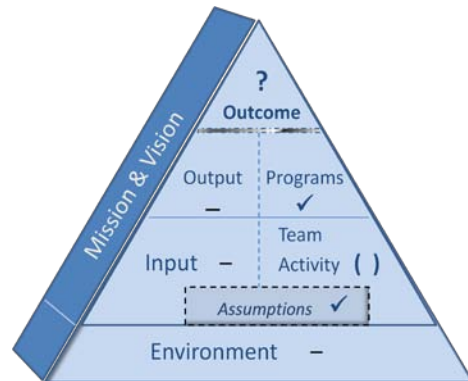

A weakness in the Team's efforts was locating useful models and standards through an external scan. A logic model could have alerted the Team of this omission. Perhaps the Team was too trusting of and committed to internal organizational structures, but we could have counter-argued that alcohol policies are contextual, regardless of conventional wisdom on the outside, as long as students, RA's, RD's, and everyone else comply with regulations.

Programming, if it could be called that, did a commendable job offering information. Numbers were available once the Team would choose to look at them – if indeed this was within the Team's domain. Causality between information and actual deterrence was uncertain. A contrary hypothesis might posit that sanctions have much stronger deterrent effect. Fines might be even more effective. Teams and task forces must conduct thorough research. I would consider recommending that Student Life accept "ownership" of evaluating this outcome.

The resulting logic model for this outcome might look quite different, with fewer elements:
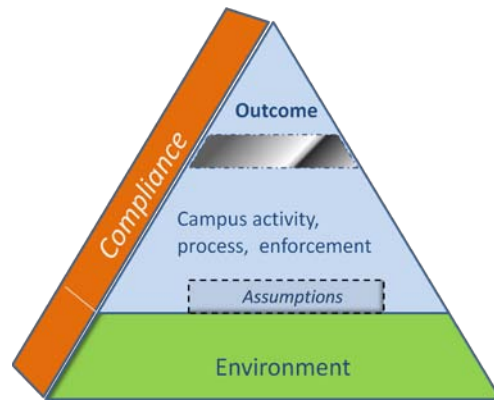
B.11   CONDUCT CODE VIOLATIONS (#10)

Outcome (Intended): "The number of repeat offenders in violation of the Student Code of Conduct will decrease."

Environment. A supplementary external scan on December 13, 2006, of "Judicial Affairs Faculty Outreach" offered results of benchmarking efforts at five aspiration institutes.

Inputs: As for the alcohol awareness outcome, inputs and characteristics were self-defined: students cited for disciplinary problems in a social setting. While demographics and background should not matter here, next phases of evaluation might consider segmenting sample populations by ethnicity and other factors. That data might indicate groups of students who are statistically at greatest risk, and thus alert planners to consider designing pre-emptive and proactive programs.

Assumptions. Originally the Team had considered that outcomes related to conduct violations, as for alcohol incidents, might best be confined to Residence Life, Student Life, and possibly Judicial Services for serious infractions (i.e., date rape, assault). A distinction from

alcohol policy violations was legality, that breaking codes were usually not enforceable off campus – merely objectionable in a social setting and subject to a hearing instead of possible trial. A breach such as cheating would tread into the academic realm, with academic repercussions as serious as expulsion. Criminal actions could result in sanctions beyond conduct hearings. In the process of evaluation, the Team would trust data from USJS and Residence Life archives to establish baselines and benchmarking.

Assessment Team activity. As for the alcohol outcome, the Team did not obtain and compile statistics for PEAR (Personal Education, Assistance, & Referral) or from University Police. A status report and "official" evaluation expected in Fall 2006 was not issued. Findings from the modest scan detailed under "environment" may have been welcome on a wider scale throughout Student Affairs beyond a memorandum to liaisons in Student Life. Further Team evaluation of this outcome was outside the time span of this study.

Participation. A critical benchmark would have been numbers of violations per student and instance, possibly with residence hall breakdowns. As USJS assimilated into Student Life, these requests lost urgency, or perhaps the analyst in charge of this program lacked confidence in available data. More data on infractions per residence hall were available in RD archives.

Outcome (Actual). Data existed but had not been processed for the time frame of this study.

Meta-evaluation.

Figure 37   Logic model depiction of evaluation of conduct code violations outcome
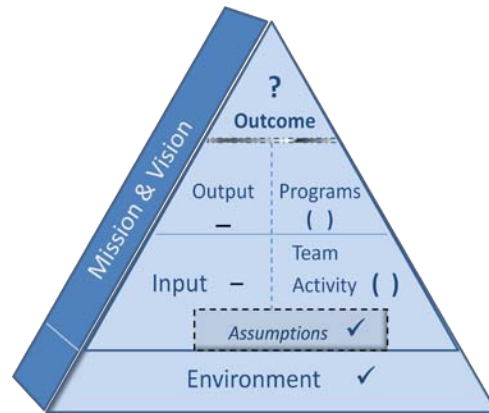
It was interesting that this outcome focused on repeat offenders. An underlying assumption was that students should have learned from their first mistake.

Summary

After a stunning rookie season, the Team's energy seemed to diminish. Other possible explanations are that we had saved for later the most difficult and contentious outcomes, the ones involving the greatest degree of bureaucracy and compartmentalization of data. If the Team's performance on the final two outcomes could be better described as anti-climactic than thorough, perhaps this observation is not unusual in many assessment campaigns. Future research of the literature and peer comparisons could determine if dwindling energy is common, and hopefully reveal lessons on how assessment teams and practitioners might maintain endurance through a campaign.

APPENDIX C

MEETING MINUTES

(This document is a sample of a typical meeting minutes recorded by co-chairs, drafted into memorandum format, and released to the Dean of Student Affairs and other interested parties. Pseudonyms preserve anonymity of Team members.)

Assessment Team Meeting Minutes
November 1, 2005
130 WPU

Attendees: [names of team members / list of nine names + Chairs, Liaisons]

Excused: [three names]

1. Since there were no requests for clarification of our previous meeting and minutes, we proceeded directly into team roles and assignments.

2. At a previous meeting, co-chairs and liaisons considered candidate programs for evaluation identified in Baseline / Benchmarking / Outcomes documents (received from all units except for International Services). Co-chairs determined first priorities as the community **mentoring** project and **electronic portfolios**. Co-chairs defined groups to focus on each project:

Electronic portfolios	Horatio Valoria Grizelda Antoinette Shaggy Scooby
Community Mentoring Project	Methuselah Delilah Giuseppe Tonto Xe-Fong

3. Individual participants and the team collectively identified "building blocks" upon which to launch our evaluations:

   a) [Tonto] provided several promising leads from an internet investigation. These findings give direction for further investigation on mentoring communities and also are a useful model for the electronic portfolio sub-committee to conduct a similar "scan".

   b) [Grizelda] shared preliminary (summative) findings from career services experience that put into doubt the value of e-portfolios to *employers* (who may lack time and energy to analyze such an expansive document). On the other hand, e-portfolios may be useful for graduate job-seekers in honing their interviewing, organizational, salesmanship and presentation skills, along with enhancing their confidence.

   Conclusion :   Continue to investigate useful models for possible pilot program at Pitt.

   Issue / framing :Are e-portfolios best constructed through an outsourced agreement and software packages, by creation of in-house expertise and technical support, or organic and independent creation by students (now the prevalent "system")? What arrangement works best and most economically here?

   c) The Team has identified existing programs at Pitt as possible foundations for a more viable mentoring community, in line with defined Student Affairs outcomes. One issue is that presently these programs represent single events, as opposed to the ongoing and process-based relationships to which a true mentoring community could aspire.

   1) Pathway to Professions
   2) Pitt Career Network (> 1100 alumni)
   3) CAS freshman mentors
   4) Greek academic mentoring
   5) Living & Learning Communities (LLC)

➡

✳   Next evaluation tasks   [responsible member] :

A. **Mentoring**
   1) Investigate another "stream" of Internet leads from the main source located by Tonto, Mentors Peer Resources (www.mentors.ca/mentorprograms.html) . [Xe-Fong]

2) Dig deeper into sources that Tonto located. [Methuslah had volunteered for this task but has closer contact with Greek and CAS programs.]
3) Continue analysis of possible pilot programs, most likely from the list above. [Methuslah]
4) Explore link to international students. [Delilah?]
5) Coordinate and collaborate with each other to compile findings into brief report for submission to co-chairs for review. [All yinz!]

B. **e-portfolios**
1) Conduct internal scan (via internet). [Shaggy]
2) Contact School of Information Sciences for precedent, models. [Scooby]
3) Contact Technology Team to share research. [Horatio]
4) Contact Ball State and other peer institutes for advice and precedent. [Chair 2]
5) Examine international templates and portfolio models. [Antoinette]
6) Coordinate and collaborate with each other to compile findings into brief report for submission to co-chairs for review. [Chair 1]

C. **Project management**
1) Collect findings; determine how all team members may contribute. [Chairs]
2) With Team input, determine next priority batch of projects; a list is forthcoming. [Chairs]
3) Coordinate with other Student Affairs teams; locate list of committee chairs' names. [Chair 2]

⇨ Our next meeting is Tuesday, December 6, from 12-2 in the Health Education office of Student Health, conference rooms B and C.

Keep in touch before then! jcc

BIBLIOGRAPHY

American Association for Higher Education (AAHE). (1992). Principles of good practice for assessing student learning. Washington, DC: AAHE.

American Association for Higher Education (AAHE) Bulletin, (1995, November), 48 (2), pp. 7-9, accessed September 2005 at http://www.uaf.edu/provost/outcomes/ .

American Evaluation Association. (2004). Guiding principles for evaluators, Fairhaven, MA, accessed at http://www.eval.org/GPTraining/GPTrainingOverview.asp .

Andres, L. & Carpenter, S. (1997, December). "Today's higher education students: Issues of admission, retention, transfer, and attrition in relation to changing student demographics," Centre for Policy Studies in Education, Vancouver: University of British Columbia.

Asian Development Bank. (2007, Dec.). Technical assistance report. Project # 40347.

Association of Fraternity Advisors Student Development Committee. Sorority and Fraternity Mentoring Communities. Accessed November 2005 at www.fraternityadvisors.org/pdf/student/dev/2005_03_Stud_Devel.pdf .

Astin, A. (1991). The assessment model. Portland: Oryx.

Ball State University, Parents Advisory Council. (2008). Education redefined ®. Accessed at http://www.bsu.edu/pac/ .

Barlow, K. (2007, January 11). "Michigan's experience could influence some Pitt research programs", *University Times*, 39:1, p. 3.

Barlow, K. (2008, May 29). "Mentors, mentees tell what works", *University Times*, 40:19, pp. 3-4.

Barlow, K. (2008, May 29). "Not your typical study-abroad trip", *University Times*, 40:19, pp. 1, 10.

Barrett, Helen C. (2001). "Electronic portfolios - A chapter in educational technology; an encyclopedia to be published by ABC-CLIO, 2001. Accessed at http://electronicportfolios.org/portfolios/encyclopediaentry.htm .

Benjamin, M. (1994). The quality of student life: Toward a coherent conceptualization. *Social Indicators Research*, 31(3), 205-264.

Bonnet, D. (1981). "Five phases of purposeful inquiry". In R. Brandt (Ed.), *Applied strategies for curriculum evaluation*. Alexandria, VA, Association for Supervision and Curriculum Development, pp. 9-24.

Bracey, G. (1997, Oct.). "The seventh Bracey report on the condition of public education", *Phi Delta Kappan*, 79 (2), pp. 120-137.

Brandt, R., Ed. (1981). *Applied strategies for curriculum evaluation*. Alexandria, VA, Association for Supervision and Curriculum Development.

Brooks, S. (2008, August 19). "Pennsylvania Hall to offer leadership living-learning community", *Pitt Chronicle*, IX:22, p. 3.

Chelimsky E. (1997). "The coming transformations in evaluation". In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 1-26.

Chelimsky, E. (1997). "The political environment of evaluation and what it means for the development of the field". In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 53-68.

Chelimsky E. (1997). "Thoughts for a new evaluation society", *Evaluation* 3(1): 97-118.

Coburn, K. L. (2006, Jul-Aug.) Organizing a ground crew for today's helicopter parents. *About Campus*, 11 (3): 9-16.

Colorado State University. (2008). Writing guides, conducting experiments: Differences between experimental and quasi-experimental research, accessed at http://writing.colostate.edu/guides/research/experiment/pop3e.cfm .

Columbia College. (2006, July). Student learning outcomes, accessed at http://columbia.yosemite.cc.ca.us/slo/student_learning_outcomes_assessment_cycle.htm .

Cook, S., Parker, R. S., Pettijohn, C.E. (2004, January/February). "The Perceptions of interns: A longitudinal case study." *Journal of Education for Business*.

Cook, T. D. (1997). "Lessons learned in evaluation over the past 25 years". In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 30-52.

Creative Research Systems. (2009). "The survey system", accessed at http://www.surveysystem.com/sscalc.htm .

Cronbach, L. & associates (1980). Toward reform of program evaluation. San Francisco, Jossey Bass.

Deepwell, F. (2002). Towards capturing complexity: an interactive framework for institutional evaluation. Centre for Higher Education Development, Coventry University, UK. Accessed at http://www.ifets.info/journals/5_3/deepwell.html .

Dillman, D. (2000). *Mail and internet surveys*. New York, John Wiley & Sons, Inc.

Driscoll, A. and de Noriega, D. C. (2006). *Taking ownership of accreditation*. Sterling, VA: Stylus.

Eisner, E. W. (1981). "Using professional judgment". In R. Brandt (Ed.), *Applied strategies for curriculum evaluation*. Alexandria, VA, Association for Supervision and Curriculum Development, pp. 41-47.

Eisner, E. W. (1998). *The enlightened eye: Qualitative inquiry and the enhancement of educational practice*, Upper Saddle River, NJ: Prentice Hall.

Elmore, R. (1996). Getting to scale with good education practice, *Harvard Education Review*, 66(1), 1-26.

Evans, N. (1996). Theories of student development. In S. Komives, D. Woodard, and associates (1996). *Student services: A handbook for the profession* (3d ed.). San Francisco, Jossey-Bass.

Evans, N., Forney, D., & Guido-DiBrito, F. (1998). *Student development in college: Theory, research and practice*. San Francisco, Jossey-Bass.

Fetterman, D. M. (1997). "Empowerment evaluation and accreditation in education". In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 381-395.

Fitzpatrick, J., Sanders, J., & Worthen, B. (2004). *Program evaluation: Alternative approaches and practical guidelines*. New York, Longman.

Frechtling, J. A. (2007) *Logic modeling methods in program evaluation (Research methods for the social sciences)*, San Francisco: Jossey-Bass / John Wiley & Sons.

Garrett, M. D. & Zabriskie, M. S. (2004). "The influence of living-learning program participation on student-faculty interaction", *The Journal of College and University Student Housing*, 33:1, pp. 38-44, acuho-i.

Gates, S. M. & others, Ed. (2002). "Ensuring quality and productivity in higher education: An analysis of assessment practices". *ASHE Higher Education Report*, 29:1, Hoboken: John Wiley & Sons.

Geist, G. (1977, May-June). "Internships for rehabilitation counselors: on the way to a professional career". *Journal of Rehabilitation*, 43(2):40.

Goodstein, L. D., Nolan, T. M., & Pfeiffer, J. W. (1993) *Applied strategic planning: How to develop a plan that really works*. New York, McGraw Hill.

Gredler, M. E. (1996). *Program evaluation*. Englewood Cliffs, NJ: Merrill.

Guba, E. & Lincoln, Y. (1981). *Effective evaluation: Improving usefulness*, San Francisco, Jossey Bass.

Guba, E. & Lincoln, Y. (1989). *Fourth generation evaluation*. Newbury Park: Sage.

Hart, P. (2008, May 15). "Faculty help sought in fighting student alcohol abuse on campus," *University Times*, 40:18, pp. 1,7.

Hartnell-Young, E. and Morriss, M. (1999). *Digital professional portfolios for change*. Arlington Heights: Skylight Professional Development.

Hoover, E. (2008, Feb.) Surveys of students challenge "helicopter parent" stereotypes. *The Chronicle of Higher Education*, 54 (21): A22 .

House, E. & Howe, K. (1999). *Values in evaluation and social research*. Thousand Oaks: Sage.

House, E., Section Editor. (2003). Introduction: Evaluation theory. In T. Kellaghan, D. Stufflebeam, L. Wingate, *International handbook of educational evaluation*, part one, perspectives (9-14). Dordrecht, Netherlands: Kluwer Academic Publishers.

Husen, T. (1987, February). "Policy impact of IEA research", *Comparative Education Review*, 31(1): 29-46.

Kellaghan, T., Stufflebeam, D. and Wingate, L. (2003). *International handbook of educational evaluation*, part one, perspectives. Dordrecht, Netherlands: Kluwer Academic Publishers.

Komives, S., Woodard, D., and associates (1996). *Student services: A handbook for the profession* (3d ed.). San Francisco, Jossey-Bass.

Komives, S., Woodard, D., and associates (2003). *Student services: A handbook for the profession* (4th ed.). San Francisco, Jossey-Bass.

Kuh, G. D. & others, Ed. (2007). "Piecing together the student success puzzle: Research, propositions, and recommendations". *ASHE Higher Education Report*, 32:5, Hoboken: John Wiley & Sons.

Lane, H. (1998). A study of quality assessment practices and attitudes within student affairs departments at Canadian universities. Education, Bowling Green State University.

Lipka, S. (2005, Dec.) State legislators as co-pilots: Some "helicopter parents" play politics to protect their children's interests on campuses. *The Chronicle of Higher Education*, 52 (17): A22-A23.

Lipka, S. (2007, Nov.) Helicopter parents help students, survey finds. *The Chronicle of Higher Education*, 54 (11): A1.

Lum, L. (2006, Nov.) Handling "helicopter parents". *Diverse Issues in Higher Education*, 23 (20): 40-43.

Mark, M. M., Henry, G. T., and Julnes, G. (2000) *Evaluation*. San Francisco: Jossey-Bass.

Martin, J. K. (2006). "R U signed on?" *About Campus*, November-December 2006.

Mentor Net: E-Mentoring Network for Diversity in Engineering and Science (for students) / participation colleges and Universities. Accessed November 2005 at http://www.mentornet.net/Partners/Campuses/currentcampuses.aspx .

Merriam, S. (1998). *Qualitative research and case study applications in education*. San Francisco, Jossey-Bass.

Michigan State University, Human Resources. (2006). M.E.N.T.O.R.S.: Centrally-Supported Mentoring Program, accessed at http://www.hr.msu.edu/HRsite/ProDev/Staff/LeaderProgs/mentors .

Mullen, C. A. (2005). *Fire & ice: Igniting and channeling passion in new qualitative researchers*. New York, Peter Lang.

National Mentoring Center. Assessing the potential of mentoring. Accessed November 2005 at www.nwrel.org/mentoring/panel.html .

Nichols, K. W. & Nichols, J. O. (2000). *The department head's guide to assessment implementation in administrative and educational support units*. Flemington, NJ: Agathon.

Nichols, J. O. & Nichols, K. W. (2005). *A road map for improvement of student learning and support services through assessment*. Flemington, NJ: Agathon.

Nuss, E. (2003). "The development of student affairs". In S. Komives, D. Woodard, and associates (2003). *Student services: A handbook for the profession* (4th ed.). San Francisco, Jossey-Bass.

Parks, Sharon D. (2000). *Big Questions, Worthy Dreams*. Chapter 9, "Mentoring communities". San Francisco: Jossey Bass.

Patton, M. Q. (1990). *Qualitative evaluation and research methods*. Newbury Park, Sage.

Patton, M. Q. (1997). *Utilization-focused evaluation*. London, Sage.

Patton, M. Q. (2002, January). Utilization-focused evaluation (U-FE) checklist.

Paulston, R., ed. (1996). *Social cartography: mapping ways of seeing social and educational change*. New York: Garland Pub.

Peer Systems Consulting Group, Inc. (2005). MENTORS PEER RESOURCES, Guide to the mentor program listings, accessed November 2005 at www.mentors.ca/mentorprograms.html .

Piantanida, M. & Garman, N. B. (1999). *The qualitative dissertation: a guide for students and faculty*. Thousand Oaks, CA, Corwin.

Popham, W. J. (1981). "The evaluator's curse". In R. Brandt (Ed.), *Applied strategies for curriculum evaluation*. Alexandria, VA, Association for Supervision and Curriculum Development, pp. 1-8.

Popham, W. J. (2008). *Transformative assessment*. Alexandria, VA: Association for Supervision and Curriculum Development.

Prato, L. (1996). "Internships: Invaluable experience or slave labor?" *Electronic Media*, 15(34). Accessed at http://weblinks1.epnet.com , February 2006.

Rodgers, R. (1990). "Recent theories and research underlying student development". In D. Creamer & associates, *College student development: Theory and practice for the 1990s* (pp. 27-79). Alexandria, VA, American College Personnel Association.

Rossi, P. H., Freeman, H. E., and Lipsey, M. W. (1999). *Evaluation: A systematic approach* (sixth edition). Thousand Oaks: Sage.

Sanders, J. R., Chair, The Joint Committee on Standards for Educational Evaluation. (1994). *The program evaluation standards: How to assess evaluations of educational programs* (2nd ed.). Thousand Oaks: Sage.

Schuh, J. & Upcraft, L. (2001). *Assessment practice in student affairs: an applications manual*. San Francisco, Jossey-Bass.

Schwandt, T. A. (1997). "Evaluation as practical hermeneutics". *Evaluation* 3(1): 69-83.

Schwandt, T. A. (2001). "Responsiveness and everyday life". In J. Greene & T. Abrams (Eds.). *Responsive evaluation, new directions for evaluation*. San Francisco, Jossey-Bass.

Scriven, M. (1981). "The Radnor evaluation derby". In R. Brandt (Ed.), *Applied strategies for curriculum evaluation*. Alexandria, VA, Association for Supervision and Curriculum Development, pp. 34-40.

Scriven, M. (1986). "Evaluation as a paradigm for educational research". In E. House (Ed.), *New directions in educational evaluation* (pp. 53-67). Philadelphia: Falmer Press.

Scriven, M. (1991). *Evaluation thesaurus*. Newbury Park: Sage.

Scriven, M. (1997). "Truth and objectivity in evaluation". In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 477-500.

Scriven, M. (2003). "Evaluation theory and metatheory". In T. Kellaghan, D. Stufflebeam, and L. Wingate, *International handbook of educational evaluation*, part one, perspectives (pp. 15-30). Dordrecht, Netherlands: Kluwer Academic Publishers.

Shadish, W., Cook, T., & Leviton, L. (1995). *Foundations of program evaluation*. Thousand Oaks: Sage.

Six Sigma. (2009). "How to determine sample size", accessed at http://www.isixsigma.com/library/content/c000709a.asp

Smith, Mark K. (2001, 2006). 'Evaluation' in the encyclopaedia of informal education, accessed at www.infed.org/biblio/b-eval.htm .

Stake, R. E. (1978). "The case study method in social inquiry." *Educational Research* 7: 5-8.

Stake, R. E. (1994). "Case studies". In N. K. Denzin & Y. S. Lincoln, *Handbook of qualitative research* (2000). Thousand Oaks, CA, Sage.

Stake, R. E. (1997). "Advocacy in evaluation: A necessary evil?" In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 470-476.

StatSoft. (2008). Electronic textbook. Accessed at http://www.statsoft.com/textbook/stathome.html?stbasic.html&1 .

Stufflebeam, D. (2003). The CIPP Model for evaluation. In T. Kellaghan, D. Stufflebeam, and L. Wingate. *International handbook of educational evaluation*, part one, perspectives (31-62). Dordrecht, Netherlands: Kluwer Academic Publishers.

Sunstein, B. & Chiseri-Strater, E. (2002). *Fieldworking* (2nd ed.). Boston, Bedford / St. Martin's.

Switzer, J. S. (2005, November). "Online mentoring in higher education: Creating a technology supported learning community." Colorado State University. Accessed at http://csltd.colostate.edu/filesanddocument/mentoring.pdf .

Tananis, C. (2000). Discursive evaluation: the journey of an "epistemorph" toward an interpretive practice and inquiry, University of Pittsburgh.

Taylor-Powell, E., & Henert, E. (2008). Developing a logic model: Teaching and training guide. Madison, WI: University of Wisconsin-Extension, Cooperative Extension, Program Development and Evaluation. Accessed at http://www.uwex.edu/ces/pdande .

Terenzini, P. & Upcraft, L. (1996). "Assessing program and service outcomes". In L. Upcraft & J. Schuh. *Assessment in student affairs: A guide for practitioners*. San Francisco, Jossey-Bass.

Theisen, G., Achola, P. & Boakari, F. (1983). "The underachievement of cross-national studies of achievement", *Comparative Education Review*, 27 (1).

Thelin, J. (2003). "Historical overview of American higher education". In S. Komives, D. Woodard, and associates, *Student services: A handbook for the profession* (4th ed.). San Francisco, Jossey-Bass.

Thrasher, J. D. (2009, February 17). "Thousands register for OCC," The *Pitt News*, 99:116, p. 1.

Timko, G. (1999). Assessing student needs for educational programming in student affairs, The Ohio State University.

Trochim, W. M. K. (2006). Research methods knowledge base: Quasi-experimental design. Ithaca, New York: Atomic Dog Publishing, accessed at http://www.socialresearchmethods.net/kb/quasiexp.php .

University of Michigan, Division of Student Affairs. Accessed 2005 at http://www.umich.edu/~ovpsa/faculty_staff.htm .

University of Michigan, International Center. (2006). Accessed at http://www.umich.edu/~icenter/ .

University of North Carolina. (2006). Student Affairs. Accessed at http://studentaffairs.unc.edu/staffdev/ .

University of Pittsburgh (March 2006), Student Health Services, accessed at http://www.studhlth.pitt.edu/healthed/pear.html .

University of Pittsburgh, Student Affairs. (June 2007). Accessed at http://www.studentaffairs.pitt.edu/ .

University of Virginia. (2008). Accessed at http://www.virginia.edu/vpsa/ .

University of Wisconsin – Extension. (2008). Program development and evaluation. Accessed at http://www.uwex.edu/ces/pdande/evaluation/evallogicmodel.html .

Upcraft, L. & Schuh, J. (1996). *Assessment in student affairs: A guide for practitioners*. San Francisco, Jossey-Bass.

Walken, R. (1986). Three good reasons for not doing case studies in curriculum research. In E. House (Ed.), *New directions in educational evaluation* (103-116). Philadelphia: Falmer Press.

Webster, W. (1981). "CIPP in local evaluation". In R. Brandt (Ed.), *Applied strategies for curriculum evaluation* (48-57). Alexandria, VA, Association for Supervision and Curriculum Development.

Weidman, J. (2006). "Socialization of students in higher education: Organizational perspectives". In C. F. Conrad & R. C. Serlin, *The SAGE handbook for research in education* (253-262). Thousand Oaks: SAGE Publications.

Weiss, C. (1997). *Evaluation*, 2nd Ed. Upper Saddle River, NJ: Prentice Hall .

Wheelan, B. (2008, July 19). "Assessment, accreditation and the political environment in which we do our work," keynote speech at Nuventive Users Conference, Pittsburgh, PA.

White, P. L. (2008, August 19). "Sharpening the competitive edge", *Pitt Chronicle*, IX:22, p. 3.

Wholey, J. S. (1997). "Trends in performance measurement: Challenges for evaluators". In E. Chelimsky & W. Shadish (Eds.), *Evaluation for the 21st century, a handbook*. Thousand Oaks: Sage, pp. 124-133.

Wikipedia, the free encyclopedia. (2009). Accessed at http://en.wikipedia.org/wiki .

W. K. Kellogg Foundation (2007). Logic model development guide, Battle Creek, MI. Accessed at http://www.wkkf.org/Default.aspx?tabid=90&CID=281&ItemID=2810002&NID=2820002&LanguageID=0 .

Wolf, Kenneth (1999). Leading the professional portfolio process for change. Arlington Heights, IL: Skylight Professional Development .

Worthen, B. R. & Sanders, J. R. (1987). *Educational evaluation: Alternative approaches and practical guidelines*. White Plains, NY: Longman.