

**Learning at Evergreen**

**II:**

**WRITING AND THINKING**

Kirk Thompson



**An Assessment Study Group Report**

**The Evergreen State College**

---

(c) 1992 by Kirk Thompson  
Revised February and June, 1993

*Additional copies available from:*

Office of Research and Planning  
3235 Evans Library  
The Evergreen State College  
Olympia, WA 98505  
(206) 866-6000, x. 6186

## CONTENTS

INTRODUCTION	1
METHOD	3
The Population: Indigenous Graduates	5
The Texts: Narrative Self-Evaluations	5
The Scales: Composition, Communication, and Cognitive Complexity	8
BY THE NUMBERS: QUANTITATIVE FINDINGS	12
The Three Scales Considered Separately	12
A Combined Index	20
A First (Quantitative) Look at What Self-Evaluations Say	24
WHAT THEY SAY: QUALITATIVE OBSERVATIONS	29
The Lower Decile and 'The Thesis Thesis'	32
The Middle Decile and 'Having Something to Say'	39
The Top Decile and Engagement of the Self	59
Endnote on Decline and High Gain	72
WHO IS ADVANTAGED, WHO AT RISK?	77
Factor Analysis: Preparation and Learning	77
Academic and Demographic Variables: GPA, Gender, Age, and Mode of Study	79
CONCLUSION	88
Summary of Recommendations	88
Deliberation and Dialogue	89
REFERENCES	91
APPENDICES	94





## INTRODUCTION

What happens, what developments occur, in Evergreen students' writing after the freshman year? This study finds that while many students grow in their ability to express themselves and to think on paper, nevertheless improvement can be described as only moderate after the freshman year. The study is based on portfolio materials, specifically on narrative self-evaluations. The data were the self-evaluations of 'indigenous' students, those who arrived as freshmen and graduated between 1986 and 1988, having attended Evergreen for all four years. These texts were rated on three different but related scales, measuring (1) basic composition skills; (2) communication skills, or rhetorical effectiveness; and (3) cognitive complexity. The ratings and a qualitative reading of the texts reveal both strengths and weaknesses in Evergreeners' habits of thought and writing.

The overall characteristics can be summarized as follows: *The strengths of high-rated evaluations are dynamic engagement of the self in educational experience and in writing, and ability to make judgments about what matters. The weakness of low-rated evaluations stems most often from serial narration of particulars, multiple and circumstantial, uncontrolled by organizing ideas or a thesis.* After Core programs, when most formal writing instruction ends, many students still have trouble expressing themselves in an organized way on paper. They still need help with whatever makes expository writing interesting and communicative. The weaknesses chronicled here show where help is needed, and the strengths indicate a plurality of goals at which to aim.

Much of the content of this paper comes from close reading of a qualitatively stratified sample of student self-evaluations - those rated in the lowest, highest, and middle deciles on a combined index comprised of the three scales mentioned above: composition, communication, and cognitive complexity. Such a reading generates confidence that the higher-rated student self-evaluations serve their authors well as transcript material. But it also suggests that a large number of students need help in order to move beyond serial narration, in the direction of learning to develop central ideas and to address issues of substance. Advanced students have more need than is usually acknowledged for learning to improve their writing, not so much technically or formally, as in learning to think clearly on paper. If student self-evaluations are good evidence, then Evergreen students probably spend too much energy 'working on their writing' stylistically and worrying about their seminar skills, but not enough time working out the substance or content of organized and well-communicated ideas.

Help with writing will not deal with the problems revealed in self-evaluation writing if it is mainly technical, concerning formalities of grammar and style; nor if it consists of meeting in groups to discuss what the participants like about one another's writing; nor if writing takes the form primarily of journals and narratives. Students especially need help to get past the 'default strategies' of bureaucratic and formulaic writing, the simplified routines for getting something on paper. They need to learn to

make decisions about what is particularly relevant and important to say. In self-evaluations, the two prominent default strategies are exhaustive listing of particulars, and quarter-by-quarter serial narration. Both of these formulas are evasive, in the way that bureaucratic writing typically avoids issues of substance (in this case, to evaluate personal achievements). The crux of the matter is *having something to say*. This entails (1) having some points to make, or in a loose sense having a 'thesis' that communicates some sense of direction in the writing to the reader; (2) addressing some substantive subject matter, content, or issue, relevant to a world larger than the subjective self and its preferences; and (3) engaging the self as a knower and inquirer, so that in the optimal case the writer has engaged in 'thinking on paper.'

This may seem a tall order, especially because it entails the viewpoint that teaching writing and teaching thinking go hand in hand. It is, however, a reasonable viewpoint, postulated in the "writing across the curriculum" approach at Evergreen and generally borne out by the present research. Writing and thinking tend to 'co-vary,' i.e., they tend to be weak or strong together and tend to measure at parallel levels on the research scales. The same students who measure high on writing tend to measure high on cognitive development. (While the measures co-vary, the results also differ significantly, indicating that they are not just several measures of the same thing.)

Beginning with this valid first assumption that writing and thinking will progress together, we have tended to assume further that after learning basic writing skills in Core programs, writing will improve in tandem with thinking when students take more advanced programs. *This link, however, proves to be weaker than assumed.* Evergreen students tend to make more progress in complexity of thinking than in quality of writing. On the average, Evergreen students write acceptably well at the end of the freshman year, and only a few write embarrassingly badly in their final self-evaluations. External raters commented that "We were impressed that most of the papers were very 'clean' (i.e., good use of standard writing conventions)" (Lederer & Hofvendahl, 1991). Evergreen students also tend, at matriculation, to employ more complex modes of thought than students at the average college or university. But this study indicates that writing does not improve much between the end of the freshman year and graduation. Regarding thinking, the news is more positive. Thinking skills improve impressively. From matriculation to graduation, they rise as much as or more than at other colleges and universities. With the broadest of brushes, the self-evaluations paint a picture of students capable of reasonably good ideation, which is not sustained quite as consistently in language, organization, and form.

## METHOD

This study observes a group of 'indigenous' graduates, i.e., students who matriculated at Evergreen as freshmen and stayed at the College until graduation. It traces their journey through college as recorded in portfolio materials - specifically, in the narrative self-evaluations which have become part of their college transcripts. The focus is how they wrote and thought, in three dimensions: (1) their mastery of the formalities of composition, (2) their rhetorical effectiveness in communicating ideas, and (3) the complexity of the ideas they were communicating. The general approach of the study is to give equal weight to quantitative and to qualitative methods: first, to see how these students progressed 'by the numbers,' according to three rating systems; and second, to let their own words paint a description of their experience with writing, thinking, and learning.

### The Population: Indigenous Graduates

The students in this study come from a population of 165 native or indigenous seniors - those who entered Evergreen as freshmen, and then graduated between 1986 and 1988. They are all of the graduates in that period who completed all of their undergraduate work at Evergreen, having transferred no credits at all from other institutions. For the part of this research dealing with cognitive complexity, the whole population was studied; for the parts about composition and rhetorical effectiveness, a sample of 80 cases.<sup>1</sup> Indigenous students were chosen, and specifically those who transferred zero units of credit from elsewhere, because they reveal the College's maximum impact on its students.

A much larger and more complicated study would be required to assess Evergreen's impact on all of its students, including those transferring in or dropping out for different durations and at different levels (lower division and upper division). Students who do not stay from matriculation to graduation are in the majority. Thus *the students in this study do not represent all Evergreen students*, because many students are transfer students, none of whom were included in this sample. They represent *graduates who arrived as freshmen*, in a time period which can be compared to the present only until important changes occur among entering students or in the College's curriculum. This must be borne in mind throughout this report: general references to all or most students, to men and women, etc., refer only to that group *within the population of indigenous graduates*, and not to the student body at large.

---

<sup>1</sup>For one scale, communication, the data was collected for another study and the number of students was 80, including only those students with a perfect record for submitting an evaluation every year, skipping none. These students were not significantly different from the larger group. The whole group of 165 has been used for the other ratings, but whenever a combined index is used, it is limited - because it averages all three scales - to these 80 cases.

The Registrar's Office provided data from transcripts on demographics (ethnicity, gender, age, and parental education), high-school background, and academic programs (units in different modes of study). This data indicated that the indigenous graduates were predominantly white (93%) and were almost equally women and men. Most were young. In student personnel administration jargon, the modal four-year student was a 'high-school direct' who entered at 18 and graduated at 22. Seventy-five percent graduated at age 25 or younger, though 25% were older. Fourteen percent were of 'non-traditional' age, graduating after age 30. Twenty-nine percent were first-generation college students, but at least one parent of 71% of the graduates had attended some college.<sup>2</sup>

The indigenous graduates were especially diverse in high-school preparation. SAT reporting was erratic ( $n = 89$ ), because Evergreen did not require it at the time, and presumably more students remembered or reported high SATs than low ones. The average self-reported SAT score was 942, at a time when the national average was just under 900. But even this weak data proves an important point: the spread of SAT scores was quite unusual, with a lower quartile from 410 to 810 and an upper quartile from 1080 to 1330. Of the indigenous *graduates* (not just entering students), the lower quartile would not have been accepted by most four-year colleges, and the upper quartile might have been accepted by elite institutions.

Similarly, there is nothing special about the mean high-school GPA, which is a solid B (3.01); but the range is again unusual: the lower quartile ranges from D to B-/C+ (1.33 to 2.56), and the upper quartile from A-/B+ to straight A (3.42 to 3.98). One would suppose from these ranges that while some students were struggling at admission with basic math and literacy, others could have been admitted to virtually any college that attracted them.

While at Evergreen, the indigenous graduates pursued a diversity of programs, a few, for example, spending all four years in coordinated studies, and others working almost exclusively in individual contracts and internships.<sup>3</sup> But the underlying average program of study came close to the pattern often considered normative or ideal: two years of coordinated study (interdisciplinary programs, usually basic or intermediate); a year of group contracts (usually more advanced and more focused on one or two disciplines); two quarters of individual contracts (either independent study or internship), and the equivalent of a quarter's work in four-unit courses.

The only way to summarize the characteristics of this population of indigenous graduates is to point out how diverse they are, as a population all of whose members graduated from the same college. Few four-year colleges or universities would consider it reasonable or desirable to admit so diverse a group, but Evergreen is deliberately different. And these are not just the students who made a start; they are the

---

<sup>2</sup>This was supposed to be an indicator of socio-economic status, since parents who have been to college tend to occupy a higher socio-economic position; but it did not work well, because records indicated only whether or not parents had received *some* college credit. Thus those parents who had taken a community college course were lumped with those who received B.A. or advanced degrees.

<sup>3</sup>Unfortunately, data compiled by the Registrar does not distinguish individual contracts from internships. All are recorded as individual contracts.

ones who stayed all four years, completed the required number of units, and graduated. The wide ranges, particularly in high-school preparation, suggest that ratings of writing and thinking will be highly variable, and that students' problems in these areas will not all be of one kind.

### The Texts: Narrative Self-Evaluations

The texts used for rating writing and thinking were students' narrative self-evaluations (SSEs). The Evergreen Registrar possesses whole vaults of these documents, because faculty and student narrative self-evaluations function at Evergreen as transcript material, in lieu of grades. At the end of every program, almost every student writes and files a document called "The Student's Own Evaluation of Personal Achievement."<sup>4</sup> These are 'portfolio materials,' favored in assessment research because they are part of what students actually *do* in college, in contrast to intrusive testing measures which do not belong to the flow of the educational process and have not proved particularly valid or reliable anyway (Council of Presidents, 1989). Portfolio materials such as term papers, written exams, and narrative evaluations are 'ecologically valid,' in the sense that they are generated in actual learning situations and show how students work on substantive academic problems. Furthermore, portfolio evaluation can be developmental, rather than distracting or intrusive like standardized testing, because preparing portfolio documents is frequently an important learning experience in its own right (Thompson, 1989).

Evergreen contains a complex 'culture of evaluation,' in which faculty evaluate students, students evaluate themselves, and students evaluate faculty. (Faculty and deans also evaluate faculty, but that is outside the scope of this paper.) A student's transcript contains, for each separate enrollment in a full-time program of study, (1) a description of the program, (2) a faculty evaluation, and (3) "The Student's Own Evaluation of Personal Achievement."<sup>5</sup> The description and the faculty evaluation are mandatory for credit. Writing a self-evaluation is a College requirement, though it need not be presented as part of the transcript for the student to earn credit. Approximately 85% percent of students in the freshman and senior years included self-evaluations in their transcripts, with lower rates in the sophomore and junior years.<sup>6</sup> The most important determinant of whether or not a student submits an evaluation is whether or not faculty members require it. In cases where the faculty member did not require it, some bias no doubt entered, with more evaluations probably received from students who could say they had done well. In this study, only the students' self-

---

<sup>4</sup>Evaluations are also written when students change programs during the year, in which cases a year's evaluations were batched together, receiving one rating on each scale for the year. Additionally, most programs - especially Core or freshman programs - go through unofficial 'in-house' evaluation procedures at the end of every quarter. These serve both as current appraisals and as *aides memoires* for the final evaluation that goes in the student's transcript.

<sup>5</sup>Most programs are full-time; the College offers only a few separate courses.

<sup>6</sup>"Approximately" because of different lengths of programs if the student remained longer than four years. The same students who submitted evaluations for the freshman year did not always submit them in the senior year, and so computations of overall gain are based on  $n = 118$ , or 72% of all students. The communication scale was based only on evaluations for which all four years were represented ( $n = 80$ ).

evaluations were read as data, not the faculty evaluations, to see what the students were saying for themselves. (Another study reviewed the portfolio-style transcripts as a whole.)

Self-evaluations are not formal academic essays. This presents some disadvantages, and counterweighing advantages. On the minus side, SSEs are usually available only at year-end, and so they could not be used as data to study changes within the freshman year - changes which may be crucial to writing development.<sup>7</sup> Second, one cannot directly compare the results of this study to other studies of writing that use classroom essays. Nor can one say for sure, from the results presented here, how well or badly Evergreen students write academic essays.<sup>8</sup>

While self-evaluations are not formal academic essays, they were in every instance in this study examples of expository prose. They go on a form headed "The Student's Own Evaluation of Personal Achievement," which sets the tasks of locating achievements of one's own and relating them to some sort of evaluative standard or value hierarchy. Surprisingly, the choice of genres was absolutely uniform: no evaluation is a poem, none is constructed out of the dramatic interplay of different voices. In intention, none is primarily an effort at story-telling, with a tragic ending or comic surprise. All try to be what Bruner (1986) calls "paradigmatic" rather than "narrative" thought. They consist of expository prose in the form of a *report*, sometimes more formal, sometimes more personal, on one's activities and (sometimes but not always) how one evaluates them. These prose samples tend to seek, though they do not always find, conceptual structures or organizing ideas. Many lapse off into serial narration when the writer does not hit on an organizing theme to hold the material together. Some (as will be seen in the section on "What They Say") simplify the task artificially by using program activities and the quarters of the year, rather than personal achievements, as the structural components of the text. In other words, many of the efforts at paradigmatic or conceptual thought do not turn out to be highly successful, judged by the rules of the game for their own genre. But all of the authors were trying to write expository prose and none has consciously undertaken story-telling, drama, or creative writing in an artful sense.

On the plus side, then, the first advantage of self-evaluations as data is that they uniformly provide examples of one genre: expository prose. They present two additional advantages. One is that they could be rated in terms of three scales of composition, communication, and cognitive complexity, whereas assigned class essays could not be.<sup>9</sup> Another is that evaluations are probably better indicators than academic essays

---

<sup>7</sup>A study of the freshman year is in the works, using the first and last essays which students wrote during the year. Though it would have made sense to study the freshman year first, this work will appear later, because data collection necessarily takes longer: self-evaluations are available immediately, from the Registrar's files, for longitudinal study, whereas essays must be collected in the future, starting from when the project was initiated.

<sup>8</sup>Tom Maddox, the College's Writing Coordinator, emphasizes this point: if students do not improve much in writing self-evaluations, that does not prove they do not improve in writing essays, because self-evaluations are not essays in the usual sense.

<sup>9</sup>Assigned essays proved, in a pilot test, to be unratable in terms of cognitive complexity, because they were too 'context dependent,' i.e., too dependent both on the

of what sort of *transferable* writing and thinking skills the student has accrued. Howard Gardner's *The Unschooled Mind* (1991) has recently focused attention on this issue: the difficulty which students typically have in exporting academic skills to non-scholastic problem-solving. The "most stunning" results come from the sciences and indicate that honors students often cannot use college knowledge to unlock simple real-world problems. This is not, however, a problem in the sciences alone; similar research results have appeared in all disciplines - social sciences, humanities, and arts - and in different countries (pp. 3-5, 143-81).

As a case in point, academic essays are a notoriously encapsulated genre, serving as negotiable currency only in the academy itself.<sup>10</sup> By looking at an academic essay, one can probably predict how good an essay the student will write in response to the next assignment, but not how the student will address the world. In comparison to assigned essays, self-evaluations are 'warm-blooded,' especially because they have recognizable audiences and purposes. The student knows she is writing immediately to her faculty member, for discussion in an evaluation conference, and in the long run to readers of college transcripts - personnel officers and admissions officers in other institutions. Furthermore, the occasion is consequential, in that one is writing not because an assignment has to be completed, but because one's own work is the subject. The assignment is consequential for the self, because it counts not only as an academic record but as a part of one's identity. In one student's words,

The evaluation system made it very clear that I was in college for my own benefit and learning, not a grade, teacher, or diploma. Education became my responsibility, my triumph or failure.

In these senses, a self-evaluation is a more focused kind of writing than an essay, with a purpose and an audience, and it probably elicits more of the kind of writing that the student will be able to offer when called upon by the world to write some expository prose.

The review panel which authorized the present study noticed one possible problem in using SSEs as data for studying writing and thinking: sometimes faculty members help students, or fellow students help one another, to edit evaluations, so that they are not 100% the student's own work. This indeed introduces some variability in the ratings, though probably in some of the ratings more than in others: editing would have most effect on composition ratings, which could easily be improved by editing;

assignment itself and on how well the student had paid attention to one particular theme, idea, or book. With so much variability, a study based on essays would have to include a large number of cases to yield any conclusions. The same pilot study indicated that issues involving the most heated emotions - specifically, the Persian Gulf War during the fighting - could not be rated, because responses were cognitively and emotionally too diverse and extreme for coding into categories.

<sup>10</sup>They address an abstract or unidentifiable audience - in one sense, only one's teachers, or one's internalized image of the teacher; or in another sense, nobody at all. The consequentiality of the act of writing is hazy, beyond the immediate goal of academic success. Similarly, relevance to the self is problematic. The self is usually either too specifically or too nebulously engaged: either one writes assigned essays for academic success *per se*, or one does so with only the vaguest sense of why the assignment matters.

less effect on communication ratings, which would change only with very extensive editing; and least effect on cognitive complexity ratings, which could not be changed much by editing (someone else would have to take charge of basic choices about structure, ideas, and voice). To get around this whole problem of students receiving editorial assistance, the present study will use average ratings for all four years, whenever possible. Faculty and peer assistance and other situational influences should cancel out when ratings are averaged.

### **The Scales: Composition, Communication, and Cognitive Complexity**

The indigenous graduates' self-evaluations were read and rated on three scales, which are attached to this paper as Appendices I-III. The first scale rated quality of *composition*, at the syntactic and rhetorical levels. For ratings on this scale, the readers were looking at (1) syntax and grammatical conventions, (2) organization of content, and (3) stylistic considerations, particularly fluidity of expression and consistency of voice. This scale was devised by the Educational Testing Service (ETS) to create a roughly normal distribution of Evergreen students, after reading 60 freshman essays (not SSEs).<sup>11</sup> The scale's four categories range from 1, "Poor (perhaps in need of remediation)," up to 4, "Very Good to Excellent." Two ETS raters devised the scale and rated all evaluations written by the 165 indigenous graduates. A full description of the standards for grammatical conventions, organization, and style at all four levels appears in Appendix I, below (Lederer & Hofvendahl, 1991).

The second scale rated quality of *communication*, or "rhetorical effectiveness," in the ETS's phrase (ETS, 1990). This scale had originally been devised to rate high-school seniors' "reflective essays" in the California Assessment Project (CAP), a statewide effort to assess student writing. CAP readers are trained to rate three kinds of essays - observational writing, autobiographical incidents, and reflective essays. While the categories overlap, Evergreen SSEs correspond best to the third category.<sup>12</sup> In the reflective essays written explicitly for rating on this scale, the California students were asked to focus on a personal experience and to reflect on its meaning. "Reflective essays," according to the ETS,

derive from the personal experience of the writer. But beyond description and narration involved in communicating that experience, reflection requires probing into what this experience can show about the writer's life in particular, and ... about the writer's ideas of life in general.

---

<sup>11</sup>The raters were given each student's set of all self-evaluations in a batch, which could have accentuated a tendency to 'read progress into' the series. The results, however, revealed so little progress that this tendency was probably minimal.

<sup>12</sup>In his initial report on these ratings, Steve Hunter (1990) wrote that "I do not claim that this definition or the criteria ... match Evergreen's intentions for the self-evaluation in an extraordinary way.... [But] the ETS definition is as close to a general notion of what the self-evaluation at Evergreen can be as is available to us through any external agency...." The criteria seem especially useful when connected with the two other scales used in this study.



The writer of a reflective essay works to see connections between experience and ideas, to test out thinking about an idea in the light of other experiences, and to arrive at new dimensions of the initial thinking (ETS, 1990).

This is very close to what is going on in Evergreen self-evaluations. A trial run indicated that the raters trained for reading CAP essays could readily apply the same rating criteria to Evergreen evaluations. The criteria center on how well a student can *relate the particular*, the personal occasion for reflection, *to the general* - to whatever personal, social, or universal meaning the student can locate in the occasion for reflection. If the particular and the general are poorly integrated, things fall apart into concrete details and platitudes. If well integrated, then the student has communicated well or has achieved "rhetorical effectiveness," in the sense of having linked vivid detail and thoughtful generalization in such a way that each illuminates the other, and the reader understands the connection. The six categories on this scale range from 1, "Minimal Evidence of Achievement," to 6, "Exceptional Achievement." The ETS drew the raters from its pool of former CAP raters, and they read evaluations by a random sample of 80 indigenous students who wrote full sets of evaluations - one for each of the four college years.<sup>13</sup>

The third scale rated *cognitive complexity*, as determined by the Measure of Intellectual Development (MID). This measure is based on the model of intellectual development devised by William G. Perry, Jr. (1970, 1981). The model was systematized as a rating instrument by Lee Knefelkamp (1974), was refined in studies at Alverno College (Mentkowski, Moeser, & Strait, 1983), and has been presented as a rating manual by William S. Moore of the Center for the Study of Intellectual Development (1982-86, 1986, 1989). Moore, who is the principal trainer of MID raters, rated all of the indigenous students' evaluations.<sup>14</sup> The original Perry model has since been critiqued and expanded by Belenky, Clinchy, Goldberger, and Tarule in *Women's Ways of Knowing* (1986). The quantitative rating system is based on the original Perry scheme, but interpretations offered in this study have been modified to take account of Belenky et al. Perry's original scheme appears below as Appendix IIIA, and the interpretive system devised by the present writer to synthesize Perry and *Women's Ways of Knowing* appears in Appendices IIIB-IIIC.

The Perry model can be thought of as carrying on the work of Piaget at the college level, by investigating what Perry called the *Forms of Intellectual and Ethical Development in the College Years* (1970). It describes and categorizes the cognitive structures used by its subjects (Harvard students of the 1950s and 1960s), and it arranges the categories as "Positions" in a helix model (Perry, 1981), according to their cognitive complexity. The Positions represent functional relationships between the knower and the known (Belenky et al., 1986, refer to them as "perspectives," and sometimes "epistemologies"). While the original Perry model consisted of nine "Positions" or stages of cognitive complexity, the higher stages are matters mostly of personality development, and so almost all MID ratings fall in Positions 2 through 5. In Position 2, students think of knowledge as information and address it dualistically as

---

<sup>13</sup>In all ratings, whenever a student stayed longer than four years, the last three quarters were counted as the senior year.

<sup>14</sup>The ratings were submitted to him in random order, so that he did not know if he was reading a freshman, etc., evaluation, unless the author made an internal reference to the year of college.

right or wrong. In Position 3, they think of themselves as generating new knowledge by using learned skills, methods, and procedures. In Position 4, they encounter relativism and wonder what knowledge is, in a world where nothing is absolute or certain. In Position 5, they are able to appreciate and construct theories and models which are contextually relevant and functional without claiming to be true-for-all-time.<sup>15</sup>

As this review of the three scales has probably suggested, they do not measure 'writing versus thinking'; rather, they present writing and thinking as a range and continuum of interrelated skills, from the grammatical and syntactic conventions measured on the composition scale to the capacity for interpreting and formulating theories measured by the cognitive complexity (MID) scale. The composition scale begins with syntactic conventions, but moves into rhetorical effectiveness. The composition scale focuses on rhetorical effectiveness, but includes the quality and convincingness of general thought. And the cognitive complexity scale, which can be said to categorize general thought in terms of complexity, also listens for the voice of the writer, and thus loops back to include the voice, style, and rhetorical effectiveness which were measured on the composition and communication scales. So the scales and the actual subjects of composition, communication, and cognitive complexity are related and overlapping, not disjointed or discrete.

Writing and thinking are conceived in this study to be interdependent for three reasons - pragmatic, theoretical, empirical. In the first place, this study could not differentiate thinking skills from writing skills *per se*, because all of the data are from self-evaluations which are written. Oral and written thinking might be distinguished by using interviews or transcripts of evaluation conferences, in comparison to written documents. But that was simply not the point of this study, which is to understand how thinking and writing develop together.

In the second place, a liberal arts and sciences approach to writing and thinking assumes that to some degree, these skills are bound up in one another - partly distinct, but still developmentally and functionally related. In the words of Joanne Kurfiss, quoted as a sort of credo in the *Washington Assessment Group Newsletter*,

... Any writing task, taken seriously, compels the writer to think through a problem to some degree.... Conversely, to think through any idea thoroughly, most of us find it almost impossible to make progress without some written record of our thoughts.... Students' intellectual development is fostered in an environment that encourages a dialectical interplay between reading, listening, speaking, and writing, all in the service of learning and thinking about what is learned... (Kurfiss, 1983; cited in Moore, 1992).

Finally, the three scales are related empirically. Above-average composition ratings were likely to occur in conjunction with above-average communication ratings in every year, freshman through senior ( $\gamma = .69$  for freshmen,  $.55$  for sophomores,  $.66$  for juniors,  $.80$  for seniors; all chi-square  $p < .01$ ). Similar significant

---

<sup>15</sup>Access to these increasingly complex positions is cumulative, so that anyone who has reached a higher-numbered position can also use the less complex 'epistemologies' whenever they are appropriate to the subject at hand. Self-evaluations normally display elements of more than one position, with the rating indicating which one was dominant in organizing the text.

positive relationships hold for composition and cognitive complexity (gamma = freshmen, .73; sophomores, .42 ; juniors, .56; seniors, .57; all chi-square  $p < .01$ ). They also hold for communication and cognitive complexity, with the exception of the junior year (gamma = freshmen, .51; sophomores, .53; seniors, .47; all chi-square  $p < .025$ ). More will be said about these relationships later, when discussing an index that combines the three scales. For now, the point is simply that the writing and thinking do demonstrably co-vary. They do not overlap completely, but when a student scores high on one, she is likely also to score high on the others. Composition, communication, and cognitive complexity are related scales, measuring overlapping outcomes.

## BY THE NUMBERS: QUANTITATIVE FINDINGS

### The Three Scales Considered Separately

The data on composition is new, but the other two scales, communication and cognitive complexity, have been used in previous studies (Hunter, 1990, and Thompson, 1991, respectively). In this section, the new findings about composition are presented first, followed by summaries of the studies of communication and cognitive development. The overall picture is that writing does not improve very much, but that cognitive complexity does improve, especially when one can look at data for the freshman year. Students continue to develop in cognitive complexity, especially as seniors; but writing tends to arrive earlier at a plateau.

#### *Composition*

Does the quality of composition improve in self-evaluations, in terms of syntax and rhetoric, measured in terms of the rating cues on grammatical conventions, organization of content, and style or voice (Appendix I)? The answer is definitely yes - but not very much. Students do learn to write better, by conventional criteria for college essay-writing, while at Evergreen. But if there is a large change, it must occur in the freshman year, before the first evaluation is written. (The Office of Research and Planning has collected early and late freshman essays, and we should see the results in the coming year.) From the end of the freshman year to the end of the senior year, improvement is undeniable, but small.

Composition ratings indicate that the average senior scores higher than the average freshman by about one-third of a point on a four-point scale. The modal freshman and the modal senior received the same rating of 3 or "Good." But the mean increased for seniors. The average freshman was rated 2.47, right between "Adequate but undistinguished" and "Good," and the average senior was rated 2.83, or nearer to "Good" than to "Adequate." The mean within-subjects gain - i.e., the average amount by which each subject improved his or her rating - is .32, or almost exactly one-third of a rating category.

The good news is that this improvement, measured by comparing the freshman and senior means, is statistically significant (with 121 cases in a paired-sample *t*-test, the one-tailed  $p = .0005$ ). Students' writing did demonstrably and verifiably improve, with odds of less than one in a thousand that this finding could be attributed to chance. We can be confident, or virtually certain, that going to Evergreen for four years tends to result in improvement in basic composition skills.

But 'tend to' is a phrase with a built-in qualification: while the average improved, not all individuals participated in the improvement. Sorting students into

those who decline (21.5%), those who stay the same (36.4%), and those who gain in composition rating (42%), one can say that gainers are the largest or modal group. Included among the 42% are 40.5% of the whole population who improve quite substantially, by one whole rating category or more. But the same sorting indicates that most students do not improve their formal composition or basic writing, from the end of the freshman year until graduation. Unfortunately, 21.5% even receive lower ratings on composition as seniors than as freshmen. And 36.4% stay the same, adding up to 58% who do not register improvement.<sup>1</sup> This is ameliorated to some extent if one asks what percentage registered a gain in *some* year - sophomore, junior, or senior. The answer is 63.2%. This might mean only that in one year, the student's faculty member was more helpful in editing; but it could also mean that students have a kind of best writing available for some occasions, in the same way they have best clothes they don't always wear. In any event, to sum up the pattern of gain, one can say it is statistically significant but small, and that the average is pulled up by a 42% minority of gainers whose substantial progress outweighs the majority of 58% who do not gain.

The cases of decline deserve more attention: nobody, presumably, expects that students write worse after college, and yet a fifth of the students did so. Some declining ratings may be attributable to imprecision in the rating system. (After all, writing worse would seem to be a hard skill to learn.) Two other explanations deserve consideration. First, the efforts of faculty to help students with their writing on evaluations are probably more intensive in freshman (Core) programs than later; and so the higher end-of-freshman-year ratings might in some instances be attributable to organized critique groups and to faculty editorial work. This might be called 'the editing factor.'<sup>2</sup> Second, students may come over time to think of evaluations as less consequential, and not put in so much effort (the 'best clothes' explanation). Many students, puzzled by their first encounter with the evaluation process, mention that they approached it as the most difficult and consequential of writing assignments. But advanced students may become nonchalant about the evaluation process, just getting it over with. Perhaps they felt more stress and therefore polished their prose more brightly as freshmen.<sup>3</sup>

Another problem is that the median gain, though statistically significant, is just not very large. Given the qualitative names of the rating categories - "Poor,"

---

<sup>1</sup>Repeated measures indicate that the freshman-senior gain is not just a regression to the mean. Freshman-sophomore and junior-senior gain are significant predictors of overall gain ( $r = .58$  for freshman-sophomore,  $.62$  for junior-senior; both  $p < .0005$ ). (Sophomore-junior gain, however, is a 'wild card,' unrelated to overall gain.)

<sup>2</sup>As explained above, because of this editing factor, the *average* composition rating will be used later in computing an overall index, because different amounts of faculty help in different years should cancel out over the four years. The editing factor matters most with regard to composition ratings, because composition is more amenable to editorial improvement than overall communication (linking the general and the particular). It is hard to imagine a faculty member editing so much as to introduce additional cognitive complexity.

<sup>3</sup>Another possibility would be that advanced students could have received lower ratings if they creatively disregarded the 'mere conventions' of essay-writing; but the present reader noticed no case of a creatively unconventional evaluation receiving a low rating.

"Adequate but undistinguished," "Good," and "Very good to excellent" - one might reasonably hope that all or most seniors would have advanced at least by one step in the system, i.e., from "Poor" to "Adequate," or "Adequate" to "Good," or "Good" to "Very good." But the mean change is way below this, at only one-third of a category.

This is admittedly a high expectation. Studies of writing at other institutions have indicated that freshman writing is often the best writing for all four years, because many colleges and universities teach writing only as a freshman course, and writing skills atrophy during the subsequent years of test-taking. But if Evergreen teaches "writing across the curriculum" and if students are, as is widely believed, doing more writing in lieu of taking formal tests, then one might expect more improvement. If comparative data from other institutions were available and were to reveal that Evergreen is doing comparatively well, this would not speak very well for higher education in general. Even if the comparative data showed Evergreen in high favor, still the findings would circle back to the point that 58% gained not at all, when self-evaluations - which are probably good examples of 'transfer' of college-learned skills to real-world situations - are used for data.

Composition ratings were significantly related to high-school preparation. For every year, freshman through senior, students receiving lower composition ratings (1 or 2) had earned, on the average, a significantly lower high-school GPA, a little under B; and those earning higher ratings (3 or 4) had earned, on the average, higher GPAs, a little above B (*t*-tests; all  $p < .02$ ). A similar and significant relationship held between lower and higher composition ratings and SAT scores in the sophomore and senior years ( $p < .02$ ). A talent, a motivational factor, or a consistent compulsiveness about composition 'belongs to' students who did well in high school, and it persists all the way to graduation.

Regarding demographics, women were moderately better at composition than men as freshmen (chi square  $p = .05$ , gamma =  $-.31$ , i.e., moderately negative for men). Pushing the data and looking at average ratings, one sees that the men's average, initially a third of a rating behind the women's (2.30 and 2.63, respectively, for freshmen), came a little closer to the women's average, year by year, until both ended up equal (at 2.82 and 2.83, respectively, for seniors). The gap seems to have closed because more men than women registered a little gain (54.9% of men, 38.6% of women), rather than a small number of men accruing big gains. But this trend falls just short of statistical significance (chi-square  $p = .07$ , gamma =  $.32$ ).

Neither ethnicity nor parental education - an indicator of socio-economic status - yielded significant results. But the relationship with age is problematic: a little bit of evidence, gathered by looking to see who tends to receive the different composition ratings of 1, 2, 3, and 4, suggests that rating levels and age were not significantly related in the freshman and sophomore years, but were related in the junior and senior years (one-way ANOVA; for ratings of 1, mean age of juniors = 35,  $p = .002$ ; seniors = 33,  $p = .005$ ). By implication, older students were initially indistinguishable from others; but as more of the others moved ahead, some older students were left behind at the starting line. A close look reveals that the relationship depended on the low-rated composition skills of a few of the oldest students. In the junior and senior years, the few lowest ratings of 1, or "Poor (perhaps in need of remediation)," were received disproportionately by students who subsequently graduated at age 30 and above. Perhaps an age-related factor of pride or embarrassment kept them from

seeking help from the faculty or Learning Resources Center. Here is another clue that if composition is a problem for indigenous students - in this case, the older ones - then it might be even more so among transfers.

These results regarding academic preparation and demographics are generally consistent with results for the other two scales, and so they will be discussed further in relation to a combined index.

In reporting their quantitative ratings for composition, the ETS raters also included a paragraph on the quality of what they had read. As mentioned above, they found the papers "clean," in the sense of making good use of standard English conventions. And they added another paragraph, which is a tip-off to the qualitative evaluation which will follow in this report. They noted that

Most writers simply made a declaration of their achievements in a course, rather than evaluating these achievements. Many considered evaluation to be a list of what they had done for the course (books read, essays written, places gone to, projects completed, etc.), rather than an analysis of ... their growth. (Lederer and Hofvendahl, 1991).

This is a clear statement of what will prove to be the most pervasive defect of Evergreen student evaluation writing: the use of a bureaucratic strategy of listing things, instead of undertaking the writing and thinking task of *evaluation*.

To sum up about composition: Using SSEs as an indicator of writing skills that are likely to transfer to 'real world' writing tasks, one can say for sure that in the area of formal composition (grammar, syntax, organization, content, style, voice), Evergreen tends to generate better writing over time, from the end of the freshman year until graduation. But the improvement is small and is not shared by everyone. Those who did not do as well in high school, males (as freshmen), and older students are a little less likely to write well, at the level of basic composition. These findings suggest that some strengthening of writing instruction in entry-level and advanced programs would be beneficial (and by inference, especially beneficial for transfer students, especially older transfers, who probably need help with writing but are not included in this study).

### **Communication**

Steve Hunter, Director of Research and Planning, presented findings based on ETS ratings of communication or "rhetorical effectiveness" to the Assessment Study Group in 1990 (Hunter, 1990). The ETS ratings apply to "reflective essays" in which the task is to present and reflect upon a particular segment of experience and to explain what meaning might be found therein. "Rhetorical effectiveness" or communication is conceived in this task as a function of the way the particular and the general are connected: do concrete details and generalizations weave together to convey meaning, or do they fall apart before the reader's eyes into circumstantial details and generalities? The key sentence, quoted previously, is that "The writer of a reflective essay works to see connections between experience and ideas, to test out thinking about an idea in the light of other experiences, and to arrive at new dimensions of the initial thinking." Again, "Whatever thought pattern emerges, the writer's reflections explore the meaning

of the occasion beyond the personal to the general." By these criteria, excessive concreteness, or sheer circumstantial narrative, receives a low rating. So does unsubstantiated generalization, especially when expressed in testimonials and moralistic cliches. Higher-rated reflective essays "move to a different level of abstraction," or move from the particular to some more general level of meaning. The word "reflective" describes this movement, from circumstantial instance to general meaning. Communication occurs when the reader understands the writer by grasping the connection between the particular experience and the general idea. The author interprets the particular to the reader, and the reader can interpret the author's text to find out what a particular experience meant to her. On reflection, and on reading many self-evaluations and the ratings assigned to them, this appears to be a well-conceived measure of communicative ability.

The original ETS assessment project for California high-school seniors specifically assigned students the task of writing "reflective essays." But Evergreen self-evaluations are quite comparable to the kind of essays the ETS assigned. In self-evaluations, the task is to reflect upon a duration of academic experience and to establish what parts of it can be set forth as "personal achievement." The student performs an act of reflective judgment and tries to communicate whatever general meanings emerged from the flow of personal experience. The raters found they could readily transfer the rating criteria (Appendix II) from one kind of essay to the other.<sup>4</sup>

Of the 165 self-evaluations of indigenous graduates, Hunter elected to work with 80 cases. He began with the 121 complete sets of four evaluations, one for each year, and from these randomly selected a sample of 80 for submission to ETS.<sup>5</sup>

Hunter found that communication or rhetorical effectiveness in Evergreen student self-evaluations does not improve significantly, from the freshman to the senior year. For each year, the average rating was what the ETS had deemed qualitatively "Adequate" in their rating criteria (3.91 for freshmen, 3.98 for sophomores, 3.83 for juniors, and 3.98 for seniors, on the six-point scale in Appendix II). His central observation was that "the tendency of students who write higher-quality self-evaluations as freshmen is to continue to do so. And conversely, if students don't develop that ability as freshmen they are not likely to pick it up in subsequent years" (p. 5). Sorting the evaluations into three groups of those who initially rated low, average, and high uncovered no significant gain trends on the part of subgroups - just lines with dips in the junior year, indicating finally that most seniors ended up at just the same level where they began as freshmen (p. 6, fig. 2).

<sup>4</sup>Because the assigned tasks were not specified in the same way, the ratings of the Evergreen documents cannot be compared directly to those on the California reflective essays. The California students are a different population, and their essays responded to a specific assignment for which the rating criteria were devised.

<sup>5</sup>Technically, it was appropriate for Hunter to use a sample and not submit all 165 cases for rating; but the decision had the effect of making subgroup comparisons difficult, because the cells for subgroups become too small to yield statistically significant results. Hunter could not have predicted that the data might be combined with other measures to generate the present study. Unfortunately, however, his decision to use 80 cases imposes a similar limit on the parts of the present study in which the communication scale is combined with others to form a combined index, because to average several scales, one must have an entry on each scale.



Overall, Hunter was able to formulate three conclusions:

1. This sample of students produced, on average, "adequate" Reflective Essays through writing self-evaluations as judged by a group of professional readers.
2. The average quality of the self-evaluations as Reflective Essays *did not improve* as this group of students progressed from [the] freshman to senior year.
3. Students who produce higher quality Reflective Essays as [freshmen] tend to do so throughout their stay at Evergreen. Those who do not produce a high quality freshman essay are unlikely to produce one of high quality as sophomores, juniors or seniors (pp. 4-6).

Hunter's study raises a pointed question about writing in the freshman year: do students simply reproduce, in college, the quality of composition which they learned to produce in high school; or do they actually change quite a bit, between matriculation and the time they write their end-of-program evaluations? If the answer to this question tends to be simply that high-school writing habits persist in college, the implication would be that Evergreen has no particularly effective way of teaching writing. If, on the other hand, considerable change occurs during the freshman year, then Evergreen already has effective methods of teaching writing in Core programs, which need only to be extended to subsequent years. Perhaps freshman gain, added to the small gain in the senior year, would prove to be statistically significant, and impressive in relation to improvement at other institutions.<sup>6</sup> This question is under study, using the first and last essays students write in the freshman year. A pilot study using a convenience sample of essays has suggested that during the freshman year, some improvement in writing ratings does occur; but that high-school background probably accounts for as much or more of the variation in ratings as does change due to freshman writing instruction.

### ***Cognitive Complexity***

The previous scales, considered alone, do not yield particularly reassuring results. In rather fortunate contrast, the study using the Perry model to measure complexity of thought (Thompson, 1991) revealed impressive gains in the freshman year, followed by a holding pattern in the sophomore and junior years, capped by a second, smaller developmental surge in the senior year. The conclusions were as follows:

The study shows that Evergreen students differ from students elsewhere in the complexity of cognitive structures which they employ at matriculation, in their development during the freshman year, and in their level of cognitive development at graduation. It provides a "value added" measure, gain in cognitive development rating from the freshman to the senior year; and it explores the magnitude and distribution of this kind of value added. It finds

---

<sup>6</sup>At many other institutions, writing not only does not improve but actually declines, because students in many disciplines do not write many papers after the freshman year.

that most gain occurs in the freshman and senior years, and that freshmen who rank lower in cognitive development are likely to gain more than those who initially rank higher. It indicates that Evergreen tends to serve different demographic categories of students equally, and that all of Evergreen's modes of study contribute about equally to cognitive development. Initially, gain in complexity of thought is fostered by interdisciplinary coordinated studies programs; but in the third year of coordinated studies there is a point of slight diminution of returns, and more gain in cognitive development seems to accrue from advanced work in the individual contract mode (independent study and internships).

Overall, the study describes success: by this measure, the College is accomplishing its mission by providing an alternative mode of collaborative, student-centered learning that works as well as, and apparently even better than, traditional educational systems to foster cognitive development (p. 1).

If this study had been based only on SSEs from the end of the freshman year to graduation, it still would have revealed a significant increase in cognitive complexity, because of the surge in the senior year. The increase would have been small, in terms of the positions and transitions on the MID or Perry scale, but large in comparison to other institutions. But in addition, it used additional data collected by the Washington Center for Undergraduate Education, consisting of MID (Measure of Intellectual Development) essays written by a sample of freshmen at the beginning and end of the freshman year (from MacGregor, 1987). These data revealed a substantial increase in cognitive complexity ratings within the freshman year, from a matriculation pretest average of 2.91 to a significantly higher end-of-year posttest average of 3.31, or an average gain of .39 (chi-square  $p < .005$ ). In qualitative terms, students tended to begin with a conception of knowledge as skills, methods, and procedures, but ended up beginning to explore more complex relativistic thinking. This change is greater than the change at the average college or university *from matriculation until graduation*: Evergreen has matched other institutions, according to this measure of cognitive development, *within the freshman year*. The pattern of development appears in Figure 1.

The lines at the left represent change within the freshman year, measured by the Washington Center study. The lines at the right pick up at about the same place, using the SSEs of indigenous graduates. The mode climbs in the junior year, and the mean or average in the senior year. Thus the modal Evergreen student ends up in the 3-4 Transition (3.67), in transition from procedural knowing to relativistic thought; and the average Evergreen student ends up in Position 4, capable of relativistic thinking (3.92).

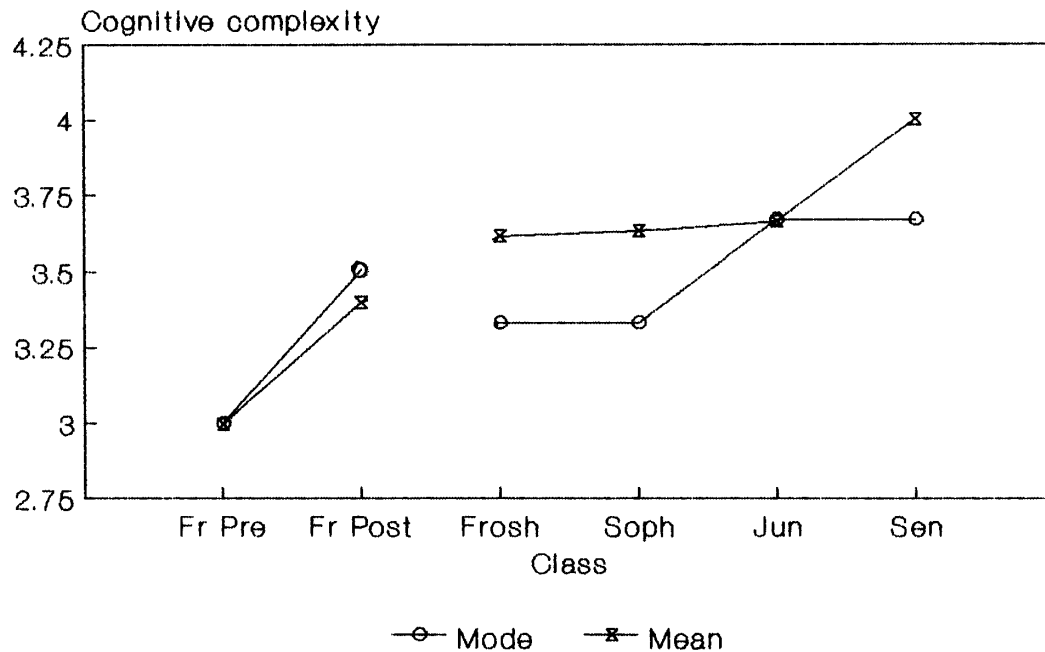
These results indicate that while the College is not so powerful in teaching formal composition or written eloquence in communication, it is teaching students to think at a level of complexity which compares favorably to other institutions.

### ***Conclusions, So Far***

After reviewing the three scales separately, some conclusions, hypotheses, and speculations are in order:

Figure 1:

FRESHMAN AND SUBSEQUENT GAIN



1. The three scales are significantly related, but do not just generate the same results three ways.
2. The composition and communication ratings tell the same story, that from the end of the freshman year through graduation, Evergreen student writing does not improve very much. The ratings are at the same time respectable and disappointing. Ratings of "adequate" to "good" composition and "adequate" communication are respectable. In grammar, style, organization, and eloquence of prose, Greeners are not exactly a rough lot, but they aren't too highly polished. On the composition side, the improvement from the end of the freshman year to graduation is significant, but not large. A good comparative sense of how large or small it is requires data from other colleges, which is not available. But in an impressionistic sense, based on the idea that one might hope every student would advance one step, from "poor" to "adequate," or "adequate" to "good," the results are disappointing. The communicative quality, or the rhetorical effectiveness, of student writing is most disappointing, because it just doesn't move.
3. More positively, Evergreen students are capable of complex thinking. One can, after all, express complex ideas somewhat awkwardly, as academic writing often

testifies. Assessment research at Washington State University has even documented a tendency of students to lose some of their formal composition toilet training when they begin to express more complex ideas and struggle for 'the words to say it' (R. Haswell, personal communication). Thus despite their lack of progress in formal composition, Greeners develop impressively in complexity of thought, both in the freshman year and overall.

4. The data about cognitive development in the freshman year are encouraging with regard to possible development in writing skills during the freshman year. One can extrapolate from the cognitive development pattern and hypothesize that when the freshman results are in on composition, we will be able to demonstrate that Evergreen teaches writing acceptably well. Freshman gains in cognitive complexity represented more than half of overall gain. While the pilot study of freshman composition has not promised gains of the same magnitude, it does foretell gains.

5. But this prediction cannot obviate the initial conclusion that writing, unlike cognitive complexity, improves little from the sophomore through the senior year. This suggests that Evergreen's current practice of emphasizing instruction in writing only in Core programs does not serve all students well. The Learning Resources Center, KEY Special Services, and some programs do, in fact, emphasize writing instruction in later years, but these endeavors are not sufficient to the magnitude of the problem. A more general approach to "writing across the curriculum," with emphasis on help for sophomores and upper-division students, would seem to be called for, if improvement in composition and communication, as measured here, are reasonable objectives of liberal arts and sciences education.

6. As a speculation, one can suppose that if sophomore through senior improvement in writing is problematic for indigenous students, then transfer students are likely to need even more help and instruction.

7. This leaves open, however, the issue of what *kind* of help, what *kind* of writing instruction. As this report proceeds, evidence will mount that students do not need technical training in writing skills, narrowly conceived; rather, they need the broadest sort of instruction in how to find words *and ideation* that express reasonable or insightful generalizations and abstractions derived from particular experiences. They must in the long run learn not just how to write, but how to think.

### **A Combined Index**

The way to connect closely with student writing is not just to tote up the ratings quantitatively, but to explore and interpret the texts. This requires sorting the cases and selecting some for close reading. This is best accomplished with a composite scale or index that combines the composition, communication, and complexity scales. This permits selecting cases for qualitative or phenomenological reading to see exactly what students say, how they express themselves, and what specific kinds of strengths and weaknesses appear in writing which was rated lower and higher. It also permits further quantitative explorations, to try to identify who is advantaged and who might be at risk.

Such an index has been compiled from the three scales, equally weighted, using each student's average for all four years on each scale. At first glance, it would make

more sense to use senior ratings, which indicate where a student ended up. On most lines of reasoning, this is what matters most. But this procedure would be biased by 'the editing factor': as the Assessment Study Group's selection panel pointed out after examining the proposal for this research, some faculty take more interest than others in helping students by editing their self-evaluations. The same point might be made about students helping one another. And so year-by-year variability might be as much a reflection of the amount of editorial assistance received as of changes in the student's own work. To deal with this, the average rating will be used in constructing the combined index, so as to cancel out variability introduced by different members of the faculty.<sup>7</sup>

This averaging procedure casts additional light on relations among the three scales. As mentioned above, the three scales are related to one another from the freshman through the senior year (with the sole exception of communication and cognitive complexity in the junior year). When the all-four-year average is computed on each scale, the scales act like linear measures - not like categories, but like numbers with decimal components. Because of this, one can now push the data a bit and examine correlations among all three scales (see Table 1, page 22).

The correlation between composition and communication is high indeed, with covariance overlap of 53%. This would justify constructing a separate writing index, which would be very accurate, from composition and communication, without including cognitive complexity. But this would not cover the full range of issues under consideration, with the three scales conceived here as overlapping scales. Thinking is already implicated in composition and especially communication. And the relationships between composition and complexity, and between communication and complexity, are substantial and almost exactly equal. Finally, the combined index will prove to be associated with demographic and academic variables as well as or better than a writing index alone. These reasons - conceptual, empirical, and pragmatic - justify constructing an overall index of writing and thinking from all three scales.

The method used to construct the overall index is to start with the average rating - averaging all four years - for each subject on each separate scale. The ratings now represent *sets* of evaluations, not individual ones. Next, the average ratings have been converted into Standard (or Z) Scores (for all students on each measure), so that the combined measure will not be affected by the different number of categories on each scale. Then the Z-scores are averaged, so that each component counts equally with the others. This produces the overall index of composition, communication, and cognitive complexity, with the three components treated equally (not weighted).<sup>8</sup>

---

<sup>7</sup>Some variability based on help from other students remains, depending on the student's willingness to receive help or insistence in seeking it out; but at some point this must be considered a legitimate part of the person's writing habits.

<sup>8</sup>An option would be to weight the scales differently, e.g., giving the two measures of writing equal weight with the one measure of thinking; or to weigh them according to the loadings derived from factor analysis. The choice here is to follow the general theme explained above, that the three scales are overlapping and to some extent 'stacked' components of liberal arts and sciences education.

**Table 1:**

**CORRELATION MATRIX:**

**COMPOSITION, COMMUNICATION, & COGNITIVE COMPLEXITY**

(Average of Four Yearly Ratings on Each Scale)

Coefficient <i>r</i> Cases 2-tailed sig.	COMPOSITION	COMMUNICATION	COMPLEXITY
COMPOSITION	1.000 ( <i>n</i> = 0) <i>p</i> = .	.7290 ( <i>n</i> = 80) <i>p</i> < .005	.5504 ( <i>n</i> = 100) <i>p</i> < .005
COMMUNICATION	.7290 ( <i>n</i> = 80) <i>p</i> < .005	1.0000 ( <i>n</i> = 0) <i>p</i> = .	.5403 ( <i>n</i> = 80) <i>p</i> < .005
COMPLEXITY	.5504 ( <i>n</i> = 100) <i>p</i> < .005	.5403 ( <i>n</i> = 80) <i>p</i> < .005	1.000 ( <i>n</i> = 0) <i>p</i> = .

Note: "." is printed if a coefficient cannot be computed.

---

The number of cases on the index is 80, which is the number of cases on the communication scale (other cases with missing values can't be averaged). The standard error of the mean for these 80 cases is .097. As Figure 2 indicates, a number of cases just below the mean fall outside the line of normal distribution, but otherwise the distribution is roughly bell-shaped. The standard deviation (an odd measure in the case of averaged standard scores) is .87. With 80 cases, this distribution can be treated as normal.

The procedures of averaging and of using Z-scores have generated a measure which is no longer ordinal (arranged in discrete serial categories) but is now interval, with an underlying measure, and fractions or decimals.<sup>9</sup> When turning later to look for quantitative relationships between other variables and the combined index, the appropriate statistical treatments will be correlation analysis with other interval variables, such as age and high-school GPA; and *t*-tests of different group means for two-valued nominal variables, such as gender and ethnicity.

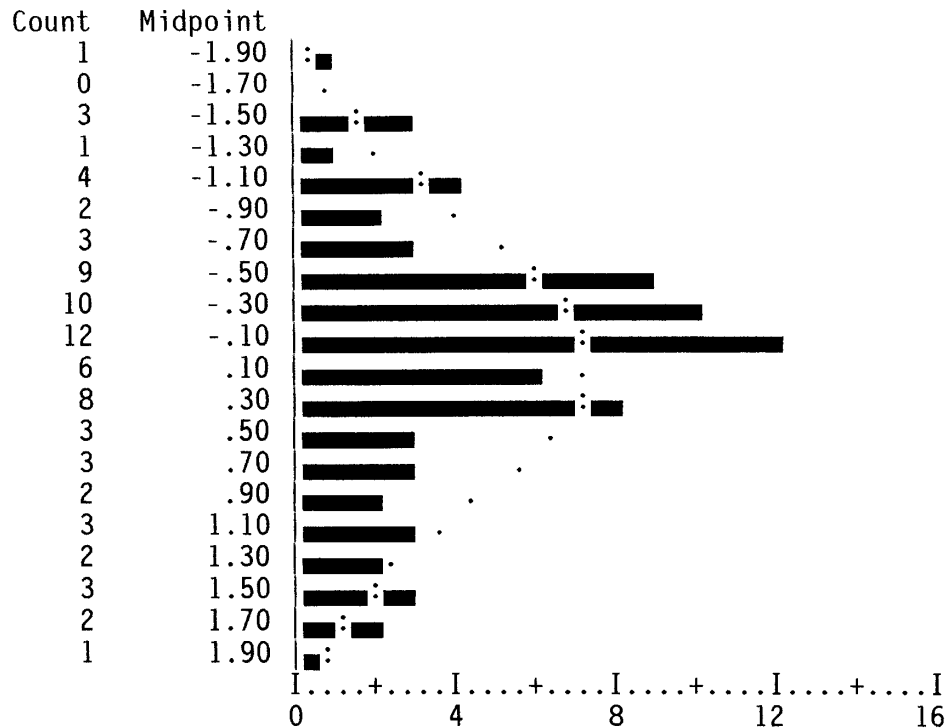
---

<sup>9</sup>The underlying unit is not something 'real,' but is the artificial result of averaging the three standard scores, which are based on standard deviations.

Figure 2:

HISTOGRAM OF COMBINED INDEX:

COMPOSITION, COMMUNICATION, AND COGNITIVE COMPLEXITY:



Note: Dotted line = normal distribution.

This index has been employed, first, to sort by deciles, or groups consisting of 10% of the sample. The index spreads out the cases, as seen in the histogram, permitting selection of the lowest-rated, average-rated, and highest-rated deciles for close reading, to see what student writing and thinking actually looks like. Reading all of the evaluations in order - as the raters and the present researcher did at the beginning of the study - produces a kind of blur in which it is hard to identify qualitative characteristics. But sorting into deciles generates contrasts and heightens differences, so that distinguishing features of overall low-, average-, and high-rated evaluations can be described and interpreted.

The same kind of computations involved in the combined index were used, secondly, to identify cases of especially high gain in writing and thinking, from the freshman to the senior year. For these purposes, the procedure was a little different,

because the comparison is between the freshman and senior years. The step of averaging all four years was omitted, and a freshman combined rating (composition, communication, and complexity) was subtracted from a senior combined rating, to pick out another 10% of the cases for close reading, to see what happened in the minority of cases where gain was considerable - those cases in which the students and the College might be deemed especially successful. This procedure was susceptible to 'the editing factor' (perhaps the cases of high gain are simply the cases which received the least editorial assistance from faculty in the freshman year and the most in the senior year); and so this is not a highly reliable procedure. Therefore it is used at the end of the next section only for supplementary information.

Finally, the index serves to establish relationships with academic and demographic variables - with the available data on high-school background, social characteristics, and modes of study at Evergreen. The point is to see if identifiable groups are advantaged or at risk, in accruing the educational benefits of formally correct and rhetorically effective writing, and complex thinking.

### **A First (Quantitative) Look at What Self-Evaluations Say**

Staying, for the moment, with quantitative findings, a preliminary impression of overall differences in the content of evaluations shows up in an informal content analysis (Table 2). The numbers in this table are topic codings, or instances of reference in self-evaluations to the themes which are listed on the left. (A full description of the themes appears in Appendix IV.) The three columns represent, respectively, the lowest, middle, and highest deciles of student self-evaluations, when these were sorted using the combined index. The number of codings by ETS raters appears opposite each theme (with references to seminars presented as a total and then subdivided into three subcategories). In general, the readers coded every time the topic changed, so that an evaluation which discussed gender awareness in one paragraph, dropped the topic, and then mentioned it again later would count for two entries. The total number of codings in the three levels of ratings appears at the bottom.

Many qualifications are in order. This is suggestive but not statistically significant data. The ETS readers were simply asked to code in the margins a series of content themes. These themes have no special status as 'the' contents of Evergreen self-evaluations. They are simply the themes which struck the present writer as frequent and significant in his preliminary reading of a sample of 100 SSEs.<sup>10</sup> The goal was only to mark passages for subsequent retrieval and rereading. Raters were neither trained nor tested for inter-rater consistency. (Indeed, some of their codings seem inconsistent with one another and with Appendix IV.) Furthermore, the topics are

---

<sup>10</sup>Specifically, these coding categories were developed for the purpose of revising the Perry model (1970, 1981) to take account of Belenky et al. (1986). Several of the coding themes represent an interest in "self, voice, and mind," as in Belenky et al.'s subtitle. The categories concerning seminars are part of an effort to track references to "voice" and "connected knowing." Issues about gender awareness, multiculturalism and racism, and also awareness of politically correct thought (either for or against it) are part of an effort to re-vision the Perry model in a contemporary context.



**Table 2:**  
**FREQUENCY OF CONTENT THEMES**  
**IN LOW-, MIDDLE-, AND HIGH-RATED EVALUATIONS**

CODES	OVERALL RATING		
	LOWEST 10%	MIDDLE 10%	HIGHEST 10%
Transition to TESC	3	3	2
Faculty and authority	0	1	11
Procedural knowledge	72	96	78
Self-awareness	18	68	61
Seminars:	25	54	38
Individual voice	(21)	(29)	(19)
Learning from peers	(3)	(12)	(11)
Seminar as process	(1)	(13)	(8)
Independent projects	5	29	14
Ideology and PC thought	0	5	10
Gender awareness	2	6	7
Multicultural awareness and racism	9	14	15
Theories, models, and metathinking	4	21	40
Commitment, agency, and calling	4	11	30
Educational theory	1	1	6
<b>TOTALS</b>	<b>143</b>	<b>309</b>	<b>312</b>

neither mutually exclusive nor exhaustive of any conception of cognitive development or self-evaluation. Nor are the codings of all evaluations in the study are tabulated here - only those from three deciles. For these reasons, this table is not a matrix for any sort of statistical analysis. It is only a rough sketch that might aid understanding without proving anything.

But after all these qualifications, Table 2 still offers a fair likeness, at least from one angle, of self-evaluations when sorted by the combined index of composition, communication, and cognitive complexity. The most obvious point is indicated by the bottom line of totals: students who rank low in composition, communication, and cognitive complexity just don't have much to say. The total number of codings is 143 for the lower 10%, versus over 300 for each of the other groups. Because the lowest 10% of evaluations are relatively short and because their content is thin, the readers found only half as many themes to encode. Presenting much less content is presumably not so much a choice as a limitation, an inability to find a 'voice' in writing, or to find 'the words to say it.'

In contrast, no difference in total number of codings distinguishes the middle and highest deciles. Students whose composition, communication, and cognitive development are only average are just as loquacious on paper as those who write and think especially clearly. Presenting a large array of different contents apparently becomes easy with a median level of writing skill and does not continue to increase. Greater writing skill and richer cognitive development may lead a student into greater depths in evaluation, but not into discussing a larger number of content themes. As later evidence will show, average evaluations tend to be based on the strategy of exhaustive enumeration. The big difference between the middle and higher deciles thus lies not in the number of points made, but in the organization connecting them.

The second most striking difference is in self-awareness. The table again reveals a difference separating the lowest 10% from the middle and higher groups, but not the middle and higher groups from each other. The difference between the lower group and the other groups is substantial, on the order of 1:3. In general, one can say that students less skilled in writing and complex thought either have less self-awareness, or have more trouble articulating it. But again, no difference separates the middle and higher levels. Self-awareness can be articulated by students who have mastered only a middling level of academic skill. Perhaps because "subjective knowing" is to some extent left behind in more complex perspectives (Belenky et al., 1986), expressions of self-awareness do not increase with further compositional, communicative, and cognitive development.

One might have expected a considerable increase in gender awareness, cultural awareness, and awareness of one's relationship to authority as students become able to handle, in thought and writing, a more complex world. But overall, references to these themes are few. In part, this can be attributed to the date of the data: the latest evaluations in the study come from 1988, and many were written in the early 1980s. For this reason, the low number of references to gender and to cultural themes do not constitute a negative commentary on the College's recent policy emphasis on multicultural awareness. (It might even be construed as evidence that the policy was needed.)

A third main point is the high number of codings in all three groups for what Belenky et al. (1986) call "procedural knowledge," the "epistemology" that conceives learning as development and application of discipline-specific skills, procedures, and methodologies. This is what one might call pre-relativistic and pre-theoretical knowing: one just uses the method and solves the problem. The Perry-Belenky model predicts that procedural knowing would represent the *whole* theory of learning and knowing in the lower-rated evaluations, but would be *subordinated* in higher-rated evaluations to more complex ways of knowing. If so, then methods and procedures would be not separate and autonomous "ways of knowing" in higher-rated evaluations, but just the intellectual tools one uses, relative to appropriate kinds of problems. But whether or not procedural knowing is later subordinated to other thought processes, the point from the table is that skill-method-procedure themes constitute a high proportion of all coded contents.<sup>11</sup>

Students in all three deciles frequently discuss what goes on in seminars. This is reasonable enough, in light of widespread belief that seminars are the heart of programs. But students exercising more limited verbal and cognitive options tend to focus only on the personal problem of finding an individual "voice" (see Belenky et al., 1986). They are less likely to discuss seminars in terms of how they have learned from peers through dialogue, or in terms of understanding the seminar process itself.

Finally, some of the coded topics are more 'theoretical': how to understand ideologies, how to understand or formulate theories and models, how to think about thinking itself, how to relate theory to practice, and how to conceive the overall aims and processes of education. Concern with these topics is expressed mostly by students in the middle and higher groups. Perhaps students rated in the lower decile have trouble entertaining such abstract concerns in the first place, or perhaps they entertain them without finding words to put them on paper. In theoretical thinking, the contrast is greatest between the middle and high groups. The task of theoretical thinking calls forth more than just moderate writing and thinking skills. In particular, the high group more often brings the issue of self-evaluation home to the self, raising questions about the author's commitment, her agential powers, or her 'calling' in life.

This informal content analysis, though weak in claims to validity and reliability, nevertheless roughly sketches some important differences among evaluations rated in different deciles. If one can trust this data to any degree, then students rated lower in composition, communication, and cognitive complexity do not have as much to say. They do not express as much self-awareness and do not often bring education home to the self, in terms of agency and commitment. They are as concerned as other students with finding a personal voice in seminars, but they are less reflective about the dialogic

---

<sup>11</sup>The study of cognitive development ended with a conclusion that most learning at Evergreen does go on at this level: "Though they move ahead of their peers at other institutions, Evergreen students still tend to be at one with most of American higher education, conceiving knowledge and learning as problem-solving and procedural knowing. The students' self-evaluations reveal that the great bulk of learning activity remains in this heartland, and though faculty may have much more complex objectives in mind, most students are perceiving what is taught within the perspective of a plurality of skills and methodologies, without yet fully appreciating the faculty's concerns with contextual relativism and more complex cognitive operations." (Thompson, 1991, pp. 79-80.)

or dialectical learning process. They are as likely as other students to write about learning skills, procedures, and methods - the 'how to' dimension of education - but they are less likely to appreciate theoretical implications. Apparently quite a bit is at stake in what the index measures, in terms of what is widely anticipated and valued in liberal arts and sciences education.

### **WHAT THEY SAY: QUALITATIVE OBSERVATIONS**

The main use of the combined quantitative index is to shift from quantities to qualities, to select cases in order to view their content and get a sense of what they say.

Self-evaluation writing is an unusual, perhaps unique, opportunity in higher education for the student to include in a transcript her own write-up of "The Student's Own Evaluation of Personal Achievement," her own version of the truth of recent experience. If formal or assigned essay-writing has a fixed form and constrains the expression of the self, then self-evaluations should prove to be different. They too are written within constraints, imposed by the imagined audience - usually a local audience of faculty and an external audience of personnel officers, graduate admissions committees, etc. - all of whom might seem intimidating to address with one's modest achievements. But the situation is a 'real life' one in which the student is asked to address the world about what she knows best: the structure, content, and above all the value ("The Student's Own E-*valu*-ation...") of her own recent experience. This is a fair place to look for her power to describe and communicate, for the structure of her thought as she makes substantive statements, for her presentation of the self in a meaning-laden context. A series of four such statements, read in order from the freshman to the senior years, should reveal the transferrable skills with which students will address the post-baccalaureate world.

In the combined index, these documents (on the average for the series) have been ranked or rated by the criteria spelled out in the appendices, for qualities or values commonly sought in student writing. The three deciles, lowest, middle, and highest, present contrasts and highlight different patterns in 'what they say.' The procedure of selecting deciles intentionally heightens contrasts by dividing up an otherwise potentially seamless continuum. This procedure creates a compare-and-contrast type of analytic problem. The possibility still remains, on turning to qualitative interpretation, that no patterns will emerge - that the three deciles will themselves be too heterogeneous to allow characterization. But the use of deciles favors finding contrasts and making generalizations about what works and doesn't work in thinking and writing.

In general, the deciles can be said to contain patterns, both of weakness and of strength. The patterns of weakness can be interpreted as areas of need, i.e., areas where the students need assistance in moving in the directions which emerge as areas of strength. This procedure has, of course, its own weakness and strength. Its weakness is the inevitable circularity: it rates what students are doing as better and worse, and then looks for further evidence within its own rating groups for ways to create more of what was initially conceived as better, and less of worse. If one dissents initially from the criteria for successful student writing embodied in the index (Appendices I-III) and

explained above, then the circularity is objectionable.<sup>1</sup> If, on the other hand, the criteria seem reasonable or even consensual regarding writing and thinking skills included in liberal arts and sciences education, then the circularity turns into a strength: one can use existing measures and already-trained raters; one can combine scales to reduce one-dimensionality and increase complexity; and one can then look at Evergreen student writing itself in order to generate 'thick description' and interpretation. The interpretation may point to deficiencies that can be improved upon and strengths that can be cultivated, once one knows for sure they're there.

A warning is in order, that the several deciles should not be stacked up mentally to build one monolithic picture of individual student development. No *Pilgrim's Progress* narrative is intended, in which Ideal Student journeys through the intellectual inferno to purgatory and heaven, or traverses all stations of the academic cross, to be reborn in literary and conceptual virtue. In only one group of evaluations, a few cases of especially high gain from the freshman to the senior year, can one see movement by single individuals from particularly low to particularly high ratings. This is empirically rare. The deciles themselves contain diversity, especially the lowest and highest. They barely reduce to three separate pictures, and definitely not to one overall picture. If what is written here has to be read as if it were a narrative, then it will have to be read as plural stories of diversities and differences, not one story of a uniform path.

These diversities and differences can be sketched very roughly as follows: The lowest-rated evaluations range from cases in which one can't see why the person graduated from college to cases in which one can see distinct improvement. When one sees the latter, it tends to be improvement in relation to a fixed initial standard, the somewhat bureaucratic standard of 'getting things right' (in contrast to a more 'developmental' change in which the student's own conception of personal achievement undergoes evolution). The low-rated evaluations often lack structure, but the more organized ones tend to employ two 'default' strategies (one or the other, or both in combination) which also dominate the average-rated evaluations. Both strategies center on serial narration of concrete details. One strategy is to say something about each of the organized activities of which the program consisted, and the other is to structure the

---

<sup>1</sup>The most objectionable form of circular rating and interpretation would occur if creative thinking and writing received low ratings because they deviated from a Standard English composition norm of the lowest-common-denominator variety. On this issue, the reader is asked simply to take the present writer's word. First, these evaluations contained no instance of a 'postmodern' evaluation which redefined the genre. Second, I found no instance in which a creative evaluation lacked appreciation by the readers, who were apparently as hungry as other teachers for originality. Third, concerned about too strong a bias toward Standard English prose, I spot-checked a case of particularly dreadful Standard English by an art student who ended up awkwardly articulating some good ideas. I found that his work was not 'trashed' by the raters, but was ranked at the top of the lower third, which was reasonable overall. Fourth, I had read the evaluations before ratings were assigned, looking for unusual and interesting themes; and many of the evaluations which were quoted in my notes from this first reading turned up for me to read again, in the upper decile. Finally, a lot of cross-checking of anomalous cases - low composition ratings coupled with high cognitive complexity, etc., - generated confidence on my part that the raters were appreciating differences rather than just legislating normality. These are, however, matters which must be taken on testimony, without means of proof.

evaluation chronologically, with separate narratives for each of the three quarters of the academic year. Both strategies make poor reading, and both avoid the task of evaluation, strictly speaking. Evaluations which are written in these ways rarely comment on the content or substance of education by mentioning ideas or interdisciplinary themes. Their statement of personal achievement is usually an implicit one, that to have remained aboard a ship on which someone else was captain and navigator was the author's principal achievement. These low-rated evaluations indicate that if (and it is a big 'if') the authors have learned to state a thesis, a point of view, or an organizing principle which the reader can follow in their essays, they have not yet transferred this learning to a writing task addressed to the real world. They suggest that many students need help with their writing after the freshman year, especially with the organizational and structural aspects of their writing.

The average-rated evaluations look better. They are not, in all their bulk for four years, very interesting, and seem to provide evidence in favor of what is called 'summative' evaluation - an overall evaluation written near to graduation, replacing some of the earlier and weaker texts, and exemplifying the *graduate's* educational experience. The average evaluations always make sense and, as the ETS raters said, are compositionally "clean." The main concern in average evaluations is with 'learning how to learn' - with skills, methods, and procedures. These evaluations suggest that few students dwell at length on the interdisciplinary themes the faculty are presumably presenting. Students instead emphasize especially the skills of seminar and writing, at the expense of discussion of content (indeed, one gets the impression that the average Greener 'majors' in seminar and writing). Each packet of four years of evaluations, if one has the patience to read them carefully, usually contains at least one significant development or change, often in the form of a breakthrough into ideation: the student suddenly comes up with something interesting to say about the subject matter (usually disciplinary, sometimes interdisciplinary). In these instances, one can see and understand a real personal achievement. But such an idea rarely serves as the organizing principle or thesis of the evaluation. The evaluations usually do not center on "personal achievement," but on recounting program activities. If experiencing such breakthroughs and getting one's writing centered on them is what writing instruction could encourage, then such instruction cannot just be technical writing training. An inherent connection obtains between organizing one's writing around a thesis or viewpoint, and conceiving a thesis or viewpoint to write about. Writing instruction, in this sense, is education in thinking, as well.

While the strength of the average evaluations lies in such breakthroughs into having something to say, the problem is that the breakthroughs tend to be buried under so many pages of chaff. Authors of average evaluations consistently overestimate the reader's interest in the details of what they have done. They tend to use the same default strategies of serial narration, aiming either at an exhaustive enumeration of particulars or at a narrative account of what went on in each of three quarters long past (a structure that is nearly always distracting to the reader, and indicates difficulty on the author's part in projecting herself into the reader's position). The numbing detail of average self-evaluations and the lack of any clear focus, particularly on personal achievement, makes one wonder just how functional they are as transcript material. Boredom is relieved, no doubt, for the personnel or admissions officer, because the self-evaluations are stacked in the transcript with the senior year at the top; and so the reader will probably just skim or quit, having reached the point of waning interest. A researcher, however, bound to plow through the whole text, cannot help proposing

summary or 'summative' evaluation (such as is practiced at Fairhaven College), in which the student, as a senior with more developed cognitive skills, writes an overview which could then replace the program-by-program and sometimes quarter-by-quarter accounts in the transcript.

The higher-rated evaluations, though still overly long, definitely serve their authors better as transcript material. Their quality is harder to describe in any one way, and the discussion will consist of exploring different qualities of excellence, none of which appears in all evaluations but all of which form parts of a model at which the teaching of writing and thinking might aim. Making an attempt at overall characterization, one might say that the 'ideal type' of the high-rated evaluation does more than just synthesize writing and ideation. It synthesizes writing and thinking *and conception of the self*. In these evaluations, the author is much more likely to think of "personal achievement" in a way that explains why knowing *matters*, why it is important to the self. The authors achieve what developmental psychologists call 'decentration,' or an overview that allows for complex and relativistic judgments not only about what one knows, but also about what is important. The reader learns how the author is involved in what she knows or is considering. The authors of these evaluations tend to adapt an overview, to plan from one program to the next, to use programs for personal goals rather than just to 'take' them as given. Graduation is more likely to be a meaningful event related to life-choices that extend beyond college. And these evaluations really *are* evaluations, in the sense that they move way beyond compiling lists of activities, and actually place values on some accomplishments and explain why they were especially important. When the knower acknowledges personal involvement and goals and a stake in knowing, then learning is no longer a kind of banking of knowledge-capital. It becomes a mode of human self-development, a way of meaning-making that shapes the self (see Kegan, 1982).<sup>2</sup> In these ways, the reader meets the authors of the higher-rated evaluations as much more developed and individuated voices and personalities. One cannot avoid the impression that this does not depend on a writing skill, but on who the authors *are*, or are becoming, through their educational involvement.

### **The Lower Decile and 'The Thesis Thesis'**

Reading the lowest-rated decile of self-evaluations is bound to arouse a range of feelings, from wonder at why the students received college degrees, to considerable sympathy with students for whom college is a difficult but productive experience. The reader almost invariably gets the feeling that the student was trying. Even when one reads between the lines, no student in this group (nor, with one exception, in the whole study) seemed to be just goofing off, leading the Evergreen equivalent of fraternity or sorority life. If such students come to Evergreen, they mask their 'real' lives effectively, or they do not graduate and therefore do not appear in this study.

Evaluations in the lowest decile display some common characteristics, but are definitely not homogeneous. Especially as one arrives at the top of the decile, one can see considerable differences: these students are learners, struggling with college but

---

<sup>2</sup>This may be said just as accurately about science students' evaluations as about those in the humanities: the highest-rated ones go far beyond a positivistic conception of science and reflect upon the responsibilities of individual scientists to the scientific community and to society.



getting somewhere important. The best descriptive strategy thus seems to be first, to look for problems fairly widespread in these evaluations; and then to look at the ones which are, relatively speaking, the success stories - those on the brink of the second decile.

### *Having Little to Say*

Informal content analysis revealed that these evaluations had relatively little to say, with the readers coding fewer instances of thematic content. Of course this means that the evaluations tended to be shorter. But having little to say is not just a matter of unanalyzable brevity. There are ways of having little to say, and if these are understood they may prove susceptible to teaching. The problem is to figure out where students who have difficulty writing and thinking are 'coming from.' From a developmental perspective, one cannot merely write off the less successful as inadequate. They are doing what seems appropriate from their own stance and perspective, and it is amenable to interpretation and understanding.

To begin with, not having much to say characteristically involves failing to see or explain relationships between the particular and the general. The low-rated evaluations indicate that good sense lies behind the ETS' practice of focusing, in rating communication, on this relationship.<sup>3</sup> The lowest-rated evaluations are full of generalizations, and also full of concrete detail, without connection between them, as in the sentence "The two books Russia and Crime and Punishment had some very interesting information inside." On the side of generality one reads, for example, that "This has been a rewarding and enlightening year with many new topics and ideas to consider." Such statements sometimes constitute the whole evaluation, as in these two instances (each a whole evaluation, quoted in its entirety):

I feel that I have gained a great deal of knowledge since enrolling in [Program X]. Many questions that I had before have now been answered. I feel that if I remain in the program (I intend to) I could become a highly knowledgeable individual about many things.

Through this learning contract I have gained direction and motivation towards the completion of my external credit document in [Field X]. I feel as though with the assistance received through this contract I will be able to create an effectively complete representation of the life experience and knowledge gained.

These texts are brief, vague, and contentless, to the point that they might be interpreted as saying nothing at all. Yet the authors had *some* conception of accomplishing something. The writers either wrote on command, without motivation, or were trying to accomplish what seemed appropriate. Inferentially, the notion of appropriateness contained in such passages is that the authors had little notion of what was called for,

---

<sup>3</sup>This reasoning is circular insofar as weakness of this relationship is the main indicator for a low rating on the communication scale. But the point to be derived from evaluations which rank low on the combined index is that problems in linking the particular and the general tend to occur in conjunction with (or co-vary with) compositional and cognitive problems.

but tried responsibly to do what was expected. The notion of evaluating one's personal achievement was alien and was not construed as a real invitation to communicate. It was probably read as 'do what is expected of you.' In this context, the author's effort apparently went into generating a simulacrum, a look-alike of statements heard, read, or uttered before which were regarded as appropriate and responsible answers to official expectations. The general intention of placating authority lies behind this kind of empty writing, but the clues are too few to identify its roots. A similar pattern emerges when the student does not seem to know what was asked for, but feels called upon to placate authority with a testimonial of acceptance. In these instances the author seems to be pledging, as if swearing on a Bible in church or court, that he or she learned, got better, or became a more 'knowledgeable' person - though no evidence appears in substantiation.<sup>4</sup> In general, one can construe such evaluations as having been written from the perspective which Belenky et al. (1986) describe as emergence from silence: the authors seem previously to have been denied a voice of their own, and are accustomed to speaking only to placate authority. Their first efforts at self-evaluation are thus presumably appropriate to their experience before Evergreen.

Some low-rated evaluations try to articulate the general in a way that remains too abstract, so that discourse walks on stilts, above the ground of common sense:

I now perceive that shifts of perspective tend to make linguistics go hand in hand with studies of all forms of human behavior. A striking feature of 20th century philosophy has been the preoccupation with the semantics of clearly expressing abstract thought patterns. The dilemma arises when one considers what was true yesterday may not be valid today as knowledge expands the concepts of the natures of man, the universe, the foundations of ethics, and accepted social practices.

Charitably, one might construe each of these sentences in a way that would make it meaningful, but assembled together they do not make meaning. Instead, they bespeak grandiosity: the writer pretends to be in control of great ideas - indeed, of the laws of thought itself - but the pretense is obvious. Still, the writer has tried to express something. Perhaps the best interpretation is that this student has heard statements that sound like this and impress her considerably. The passage is another simulacrum, but with more excitement about the life of the mind. A kind of mimicry has taken over, but it testifies to legitimate intellectual challenge, in the form of admiration for the authority and wisdom of people who can make statements of the kind that has been simulated.

The other side of the disjunction between the general and the particular is discourse which is mired in the concrete and circumstantial:

All the classes were very interesting: for example, when Dr. [X] described the steps when using the memory function on a calculator, and his use of inclined planes, marbles, and friction cars to demonstrate the attractive force of gravity, it helped my comprehension.

...

---

<sup>4</sup>Testimonials continue to appear, but with efforts at substantiation, in the average-rated evaluations.

However, the main interest for me was learning how to measure and figure geometrically the precise way of laying floor tile.

These tendencies toward the concrete and circumstantial also take the form of list-making, sometimes without telling the reader how the categories are constituted, so that one feels one is reading Borges' famous passage about categories of animals (cited in Lakoff, 1987).

It is harder, in evaluations which focus on the concrete and the circumstantial, to infer the author's purpose. Perhaps it is just to write something and get the job over. A more charitable interpretation is that the author is addressing the problem of writing an evaluation by the reasonable means of conducting a memory search and asking what has stuck there. She then reports on this accurately and honestly, without embarrassment that someone else might call it trivial. This is a forthright and understandable procedure, though a gear has slipped, disengaging the words "Evaluation" and "Personal Achievement."

In this respect, the crucial word in the previous quotations is "interesting." One reads over and over again such phrases as "captured my interest most of all...." This phrasing is prominent in both low-rated and average-rated evaluations. The authors of these phrases are confused or avoidant when confronted by the words "Evaluation" and "Personal Achievement," and they systematically substitute the concept of personal *interest*, thus reducing the evaluative task to discussing what they liked or disliked. A kind of slippage has occurred, resulting in a category error, or a different script or scenario. The student hears not the intended question about evaluating academic achievements, but another one, frequently asked about entertainment experiences: "How did you like it?" The question has become one of taste. At worst, one might infer that education has been trivialized into a passive entertainment activity like watching television. As a viewer of television, everyone learns to exercise taste, or anyway a sort of taste, articulated in terms of likes, dislikes, interests, and enjoyment. The model of evaluation (and of choice of academic programs) would thus be channel choice, or flipping one's attention button from one channel to another until one's interest is held long enough to allow the drama to unfold. In this script or metaphor, one simply says afterwards what one liked, enjoyed, or was interested in.

A more psychological view of the self who likes, dislikes, and expresses interests appears in Robert Kegan's *The Evolving Self* (1982) as "the interpersonal self." This is a stage of development associated with the period of adolescence in which young people learn that they can 'relate to people' and trade opinions with them in conversation, thus getting some sense of who they are. In this stage, the main mode of differentiation which acts as a surrogate for identity is peer-shared taste, i.e., what one enjoys, is interested in, and likes or dislikes. (The best contemporary example is musical taste.) One's tastes are, or to the speaker seem to be, the self. Of course these taste-opinions are mass-produced, replicated almost mechanically within a class and group, and so they offer little clue to personal identity; but to the adolescent they are nonetheless a way of making meaning. One hears this in Evergreen seminars and is probably reading about it in many evaluations which turn "personal achievement" into "what interested me."

Lower-rated evaluations (and also many average-rated evaluations) sometimes contain a similar translation of the task, so that the author talks mostly about how she

"feels." The result is a generalized testimonial about the self, one that forces what one learned or studied out of the frame:

I feel that I have achieved a great deal as a student and as an individual. I am much more knowledgeable and better informed than ever before. I have restored my confidence and feel that if given a task, I could accomplish it without any problems. I am now in the right frame of mind for achieving. I intend to continue headed in that same direction.

A pedagogue might criticize this paragraph (which is an entire self-evaluation) on grounds that the author is oblivious to what education is about: she thinks academic achievement is not a kind of intellectual activity, or even a set of tasks performed, but only a feeling-state, a version of positive thinking or self-esteem. But the paragraph can be viewed more sympathetically, as saying that self-confidence *was* the central personal achievement. The main problem the student faced was one of restoring confidence in the light of previous life experience or academic experience, and of coming to think of the self as one who *can* accomplish tasks. Four years of work by this student reveals progress, if not a complete success story: though she consistently moves ahead from her own baseline, much of her accomplishment seems to recapitulate high-school work, and it is hard for a reader of the self-evaluations to agree that her level of intellectual activity at the end of the senior year would usually be rewarded with the bachelor's degree.

The evaluations quoted so far illustrate different ways of having little to say, by offering only generalities, or only concreteness; by listing a few "interests"; or by testifying briefly as to how one feels. The picture perhaps looks discouraging, making one wonder why the authors graduated. But so far, the illustrations have been drawn from a few of the lowest-rated evaluations, where the quality of the whole educational enterprise seems to have been in question. But the lowest decile is not homogeneous, and even within it, the picture improves considerably.

### ***Catching On: Some Modest Successes***

As one moves through the low-rated evaluations and approaches the second decile, the writing displays basic competence in handling the English language, in listing and describing activities, and in applying methods and procedures. As the ETS raters noted, these are not evaluations *per se*, but are mostly lists of activities included in the program, especially those which seemed "interesting." Still, the reader is likely to sense that the authors have profited from attending an institution of higher learning.

The evaluations begin to have some content. Some begin with statements of goals (even if the evaluation is not organized in terms of them). Others describe development of study habits - procedures and disciplines for getting some work done. In general, the more successful of the low-rated evaluations begin to display the way of knowing which is typical of average evaluations, the way which Belenky et al. (1986) call "separate procedural" knowing. The authors are, in Evergreen parlance, "learning how to learn":

My goals for winter quarter were to develop study habits and the discipline needed for successful completion of the work. At first it was difficult to

remember what I read. This has become easier as I have gone along. I gradually developed a routine at home which allowed for uninterrupted study time. I now feel that I can handle the workload with no problem.

As in the previous example, the reader learns how energy was expended, and can conjure in the mind's eye images of students studying:

I spent a lot of time in the library researching information, reading books and magazines that pertained to the various subjects of the course. Additionally, I asked questions during and after class that I wasn't completely clear on.

One also sees, in these evaluations, some important instances of content - instances of complex thought that bespeaks real learning. The evaluation is usually disorganized and the ideation is not sustained, but the insights are there:

What impressed me most was how all living organisms are dependent on each other in some way, whether at the beginning or end of the food chain. We all need each other in order to live.

In the midst of a set of evaluations that was rated low overall, these two thoughtful and insightful paragraphs stand out like stars, expressing something like systems theory and complex relativism:

In this program I learned to question the "normal" and "usual" way of doing things. In the criminal justice system there is room for many new ideas. I learned that everything affects everything, meaning you can't look at one thing and fix it, when many different things effect that one thing. For example, the criminal justice system is not a problem by itself. Overpopulation, our definition of mental illness, etc. all contribute to the justice system.

I learned we perpetuate mental illness to keep psychologists employed. That many people are socially ill, not mentally ill. Meaning people who don't feel they belong. I realize there is a need for people to discuss the issues in their life and receive support, but we've taken it too far. I learned that what you believe is what you see, therefore you see what you think. I want to break that and "see" and "think" many different things. To give different ideas and not always do the norm.

These paragraphs were isolated, but they exemplify styles of thought found in evaluations receiving higher than average ratings, and they represent an enormous breakthrough in the senior year. Certainly they tell a success story, in terms of the student's own baseline or starting-point.<sup>5</sup> This instance proves that even within the lowest decile, great gains were possible - though in no other instance to this degree.

---

<sup>5</sup>In isolation, these quotations may even suggest that an otherwise competent student was troubled only by a technical writing problem (distinguishing sentences from sentence fragments). Indeed, without this problem, these evaluations would have escaped the lowest decile. But they appeared among other evaluations by the same author which justified the low ratings. The construction of the index made it very hard for evaluations to end up in the lowest decile for technical reasons alone.

## Conclusions

Looking over the evaluations in the whole lowest decile, however, one has to say the norms were either no progress, or quite modest improvement. The overall problem is an excess of concreteness in writing and, presumably, thought, without structure or generalization. The evaluations tend to begin without introductions or with empty global statements on the order of "This has been a rewarding and enlightening year with many new topics and ideas to consider." They usually proceed to descriptions of particulars listed activity-by-activity or quarter-by-quarter, without actually discussing the "new topics and ideas." The general and the particular are usually disjunctive, though the descriptions of particulars tend to get "cleaner," as the ETS put it, and more extensive. But to remain in the lowest decile, either gains were extremely modest, or the baseline was so low that the student, while improving, could still retain a place near the bottom.

Overall, the center does not hold: practically none of these evaluations (besides the one just quoted) contains a thesis, theme, or organizing idea. One might call this the 'thesis thesis': the present study finds that the lowest-rated evaluations make very poor reading because of the lack of an organizing dimension. Hence the thesis that good writing needs a thesis. Evaluations composed without structure communicate little and convey little thought, simple or complex. In these respects they are perhaps fair and accurate in conveying the impression of little or no personal achievement.

Considered at the level of writing, one could help these writers by conveying the idea, not just didactically but through organized practice, that *good writing has a thesis*. If the term 'thesis' seems too formal and seems literally to apply only to an academic essay, then one can at least say that *good writing contains organizing principles*, or some threads for the reader to follow so that paragraphs do not succeed one another at random. Absolutely minimally, *good writing contains an authorial point or points of view*. Granted, nobody said that "The Student's Own Evaluation of Personal Achievement" had to be a formal essay or even expository prose. The initial point was that self-evaluations might exemplify transferable writing skills precisely because they are *not* formal academic essays. But a thesis or multiple principles of organization belong to all expository prose, including non-academic expository prose. The experience of reading self-evaluations is utterly convincing evidence of how the writer loses the reader when such structural cues are missing. And the self-evaluations *are* expository prose. None of these evaluations was conceived as a poem, drama, or story. All authors were *trying* to write expository prose, even in the least successful examples quoted at the beginning of this section. So in recommending that the texts be more organized in order to be more readable, one is not just imposing essay standards on writings in another genre, but is only suggesting that the writing be improved by being truer to itself within the author's own choice of genres. It would seem, then, that students might achieve more in the way of transferable writing and thinking skills if they were helped to organize their expository prose around theses or organizing ideas.

This leaves open the question of *how* one arrives at a thesis, which will be addressed after looking at the average-rated evaluations. But two more questions should be addressed before concluding this section: How well do these evaluations serve the purposes of a transcript? And how well does the College serve its less successful students?

One cannot deny that the lowest-rated self-evaluations present their authors fairly when included as transcript material. Those at the very bottom of the pile do not present the College well, if readers would be inclined to ask why the person graduated. But most evaluations in this decile do not raise that question, and do reveal the strengths and weaknesses of the authors, if the reader is willing to continue reading very far. Almost by definition, they report accurately and exemplify what their authors have accomplished. If one sees weaknesses in composition, communication, and cognitive complexity, then one simply knows how modest the accomplishments were. In this way, the lower-rated self-evaluations act like a grade of C or below on a conventional transcript.

One can, however, reformulate the question: Do the lowest-rated evaluations show the College and its students relating to one another optimally? Here, at least part of the answer is no. In particular, these self-evaluations (and most of the average ones) have a disturbingly non-developmental, non-problem-solving quality. Students do not appear to work on their own deficiencies, and in only a few instances do they mention such efforts in their lists of activities. In general, problems about writing and thinking tend to persist without students practicing any sort of tracking or accountability in dealing with them. One might see a fuller picture by reading faculty evaluations. And perhaps students are aware of their problems and really are trying to deal with them, but do not want to discuss this darker side of their work in their transcript. But overall, a reader of the lower-rated self-evaluations receives the impression of little conscious problem-solving with regard to writing problems. In contrast to the authors of average-rated evaluations, who often describe themselves as "working on my writing," these students tend to practice avoidance either in facing their writing problems or in discussing the ways they are working on them.

In this way, the lower-rated evaluations are not just accurate reports on modest accomplishments, but are evidence of need for help with writing. These texts are the work of some of the College's least prepared or least motivated or least talented students, in terms of what the index measures; but all have ended up graduating. They and their transferrable skills now represent the College in the community. Unless one is arrogantly sure that instruction was adequate and these students were permanently defective (stupid, unmotivated, incorrigible), then it would seem that some effort should be expended in the future to help those like them. Since all but a few of these self-evaluations were written at the end of the freshman year or later, it would seem reasonable that some of this effort be expended after the freshman year.

### **The Middle Decile and 'Having Something to Say'**

The self-evaluations rated in the middle decile on the combined index, though again heterogeneous, are still subject to some overall characterization. The impression created by the whole group is of *competence*: an ability to get ahold of what is to be done, to do it, and then to write about it fairly well. At the least - and for most of these students, most of the time - this means a kind of bureaucratic competence, achieved by putting in time and effort, taking a series of programs, meeting external expectations, and enumerating one's activities in one's self-evaluations in considerable detail. The kind of learning and knowing encountered most often in these evaluations is what Belenky et al. (1986) characterize as "procedural knowing," or knowledge at

the level of skills, methods, and procedures. In many cases, this means students have been "working on their seminar-ing" or "working on their writing" as ways of learning how to learn. These two statements appear so often that they seem even to obscure or prevent emergence of more definite statements of intellectual achievement. But the average evaluations contain, in almost every set, reassuring signs that the student can do what the world will ask, in normal and modest employment. The procedural knowing in these documents is usually procedural knowing fairly well done. Also, almost every set of average-rated evaluations contains at least one 'breakthrough into ideation' - an idea which occurred to the author and which evokes the reader's interest by virtue of its content or import. Though these ideas appear intermittently and are usually not the central point or organizing idea of the evaluation, at least they appear for the careful reader to see.

To characterize, on the other hand, the main limitation in average evaluations, one can say that both writing and thinking are constrained by severe limits on what developmental psychologists call 'decentration.' All the details are there, but the big picture is missing. These students do not stand back in intellectual space and take the long view. They tend to present a close-up view which often strikes the reader as too detailed, too trivial and boring to read. The authors do not often engage in evaluation *per se* - in figuring out which of the year's events really were personal achievements, and which of these were most important. They dwell way too extensively on matters of small consequence, especially on quarter-by-quarter reviews of routine program activities. The authors do not seem to review and revise their own writing in a way that allows their own central ideas or organizing themes to emerge. At another level, they do not step back to take a look at the overall shape or configuration of their academic programs. Few take an overview of how the program was constructed or what the faculty might have had in mind in organizing and sequencing the activities which the student describes in serial form. These students have been *in* their programs, but seem to have followed them with orientations limited to the present - like the old-fashioned kind of sing-along movie in which one follows the bouncing ball, without having to know the song or to hold the larger configuration in memory. This section will conclude by showing that the authors of average evaluations are capable of many flashes of insight into the content of their experience; but at the same time, they consistently fail to step back far enough to make their own map of the terrain they have crossed. This will lead to a recommendation not about teaching writing and thinking, but about the evaluation process itself - that the College might consider the case for 'summative' evaluations.

### ***The Average Student as Bureaucrat***

The average-rated students' evaluations, while boring to read in large numbers, were individually neither inspiring nor distressing. The writers appear reasonably competent, though the quality of performance tended to be more bureaucratic than intellectual: these students, literate and employable, tend to do everything asked of them, and to do it fairly well and diligently, without admitting to much excitement about the life of the mind.

This situation presents a problem for whoever would analyze average self-evaluations: what should be the mode, in this study, of presenting a body of writing which is characteristically boring to read, and interesting only when the authors make rela-



tively rare breakthroughs? Literary critics have identified a 'fallacy of imitative form' - the idea that one must write romantically about romances, heroically about epics, etc. Trapped in this fallacy, one would have to write boringly about boring texts. To avoid this bind, this section is less empirical, at least in terms of quoting extensively from the data. It deliberately generalizes, sketching an overall picture. The standards for empirical evidence and illustration are perhaps high enough elsewhere in the study to justify this departure. What follows is thus a sort of generalized view of the landscape from the railroad car window, during a trip across the great Midwestern plain of self-evaluations.

In content, the average students' evaluations tended to be similar to the best of the lower-rated evaluations discussed previously; but overall, these writers have much more to say. Average evaluations contain clear narrative writing which recounts program or contract activities. The account is usually circumstantial, without much organization, generalization or abstraction. It discusses involvement with particular components, tools, methods, techniques, and projects, without much decentration or capacity to take an overview. Overall, there is nothing 'wrong' with this sort of account, in terms of writing or cognition, but it tends to be long-winded and not very engaging.

At bottom, most average evaluations were neither evaluations nor statements of personal achievement *per se*. They are weak from the standpoint of evaluation, which still tends to be phrased in terms of likes, dislikes, and interests. They tended to say that the individual had participated as one member in the process of the group, and had done what was expected, as distinct from having achieved something personally. Expectations come not from the self but from the program or the faculty or Authority - from somewhere else, up the hierarchy. In this respect they were no doubt true reports. One might best characterize these documents as 'periodic accountability reports,' submitted in accord with a bureaucratic requirement, chronicling how the individual has expended energies. They rarely mention discovery or excitement. They suggest that many or most of the activities on which they report had taken on the character of routines. Thus just as the bureaucrat, going through the daily routine, is earning a salary, so is the middling student, by accruing units or serving what the Veterans' Administration calls "seat time," earning a degree.

A reader of average evaluations is bound to conclude that Evergreen students are less unusual or radical a breed than their reputation claims. The evaluations give the impression of a modern, moderately competent blandness of academic orientation and life orientation. Programs succeed one another, life goes by, and writing becomes more lucid, but not much more interesting. These evaluations lack drama, or any sense of culmination. While programs and activities are often called "interesting," the accounts do not actually arouse reader interest, because they usually do not specify why the experiences were significant, what illumination they brought, or what purpose they served. The authors experience difficulty or lack interest in achieving an overview or vantage point outside the field of current activities from which to interpret or "make meaning" (Kegan, 1982). They see things from up close, and are not even very issue-oriented: again like bureaucrats, they accept a great deal as given and are not motivated to discuss many fundamental assumptions, issues, or problems with which they have wrestled during the academic year.

This overall impression, that the average way of going to college, even at Evergreen, is a lot like working in a bureaucracy, will become clearer in retrospect, in the light of contrasts to material which will appear in later sections of this study - first, to the breakthroughs into interesting ideation, discussed in the next section; and second, to the consistently engaged and engaging evaluations rated in the highest decile.

Though the method is speculative, one can infer a view of college life which seems to underlie many average evaluations. The general impression is of *life in an institution*, as if college were still like high school, or were like any other office, army, or prison system in which one has to serve four years' time. This style of experiencing the world no doubt extends to other aspects of college life besides writing evaluations. The student simply chooses among the programs offered, perhaps grumping a little about trouble finding an interesting program, but never creating an overall academic plan. Within programs, the syllabi tell one what to do. One's accomplishment consists simply in moving through the assignments, more or less carefully reading most of the books and writing most of the papers. It does not consist in connecting the large patterns of the program or figuring out why the faculty shaped it that way. Seminars are rituals performed regularly, and they work well when one finds a voice in which to say something - to take the requisite turns to keep the ball moving, toward no special goal besides its own perpetual motion. Participation in such seminars is a kind of pseudo-work, since one can space out or remain silent with no real consequences. Faculty seem to serve a formal function ('facilitation'), or else they get in the way of what students were doing well enough on their own. Faculty aren't people with whom one makes intellectual connection. In this bureaucratic mode, education is only a ritual in which the believers must offer timely sacraments. Self-evaluations are one instance. What matters is their timeliness and extensiveness, not "personal achievement," which was not actually conceived to be the goal.

One might ask how average students *learn* to be educational bureaucrats. Certainly the lessons are taught by the surrounding culture and by the K-12 system: every student who arrives at college has indeed served time, even in the best of schools, without too much stimulation, for 12 years.

On the one hand, Evergreen encourages departures from this pattern. It attracts students who want to break the mold. It offers substantial opportunities to do so, as some passages in the average evaluations and virtually all of the high-rated evaluations testify. From this standpoint, the institution and its faculty seem to be fighting against what Freire (1970) termed the "banking" conception of education, fighting in favor of student consciousness-raising. But many average students are not attending this alternative college. They have brought another one with them, in their minds. Raised in the K-12 mold, they slip back into habits of serving time.

But on the other hand, whenever an attitude is projected, there must have been a hook to hang it on. Some aspects of Evergreen must be undermining its own alternative. The college must present as *one* of its countenances a facet - not just of its administration, but of its academic side - which average students can read or interpret as bureaucracy, if prone to do so. Even an alternative institution tends to become bureaucratized, as its priests and priestesses repeat through time the rituals which once inspired, and forget just a little more of the why, the wherefore, and the spirit with each repetition. Weber (1958) called this "the routinization of charisma"; and perhaps this is what average students partly project and partly perceive.

This is admittedly a speculative characterization of 'the average student as bureaucrat.' More specific evidence appears in three aspects of the evaluations: (1) a passion for exhaustive enumeration of particulars; (2) an inclination to use the 'default strategies' of organizing one's account of intellectual life as a list of program activities, often quarter-by-quarter; and (3) a tendency to settle for what Belenky et al. (1986) call "procedural knowing" - the sort of "epistemology" in which the student is attracted to methods and procedures like "seminaring" or "working on my writing," without yet being able to reach further toward evaluative, deliberative, or creative thinking (in Belenky et al.'s vocabulary, "constructed knowing").

### ***Exhaustive Enumeration***

It would be painful to quote, or even to measure quantitatively by inches or pages, the space in average students' self-evaluations devoted to enumerating and listing activities, without focus on personal achievement or program themes. Why might a student construct the inclination (even the *need*) to enumerate exhaustive details of collective experience? Is this a defense against evaluation *per se*? Or a reasonable practice, conceived from a point of view in which quantity of collective activities seems identical to quality of personal achievements?

While lower-rated evaluations tended toward brevity, the average ones are a tall stack, as tall as the highest-rated ones. These students have found the words to say it, usually by recourse to the bureaucratic method of exhaustive enumeration. One might say the middle-decile evaluation form has been retitled, so that it no longer reads "The Student's Own Evaluation of Personal Achievement," but "The Student's Own Exhaustive Enumeration of Program Activities." An evaluation of personal achievements might seem, in an old-fashioned logical way, to require some categorizing and prioritizing, first in terms of what was personal, as distinct from the activities of the group; and second in terms of what was an achievement, as distinct from a routine. But either this is not so, or the authors did not perceive it, or they perceived it but ran a defensive maneuver in another direction. Exhaustive enumeration indeed points a different way: enumeration replaces evaluation; exhaustiveness eliminates priorities; and group activities and routines take up all the space that might have been devoted to more uniquely personal achievement.

One might claim, on the one hand, that there are no rules of the game of self-evaluation, and so whatever the students do *is* the game. If students did not do what one might have expected, then the fault lies with the expectations. A student may conceive her personal achievements in any way she sees fit. But on the other hand, not all conceptions are equally functional when external readers of transcripts enter the picture, and particularly when the word "evaluation" confronts the reader at the top of the page. The reader forms an expectation that what follows will *be* an evaluation, and if it is not, then the content seems confused or confusing in terms of its own heading. "Evaluation" and "Personal Achievement" have commonsense implications. Evaluation, in common usage, means exploring the value or worth of something. It implies creation of some sort of value structure in which some things or experiences are more important than others, i.e., constitute greater and lesser "achievements." This difference between enumerative listing and evaluation was clear to the ETS readers, cited before as saying that

Most writers simply made a declaration of their achievements in a course, rather than evaluating those achievements. Many considered evaluation to be a list of what they had done for the course (books read, essays written, places gone to, projects completed, etc.), rather than an analysis of the course or their growth (Lederer & Hofvendahl, 1991).

From this standpoint, the transformation of the task from evaluation to enumeration is a substantial change, and one might conclude that exhaustive enumeration has been substituted as a *defense against* evaluation. The defense might be conscious or unconscious. Either way, one could infer that average students don't like to evaluate themselves and are using exhaustive enumeration to derail or short-circuit the process.

But this is a speculation, not a point that can be settled by the data. What is clear is that in enumerating everything that happened, the authors consistently, and even vastly, overestimate the interest of readers in the concrete details of their academic life (assuming that the readers include admissions and personnel officers and Evergreen faculty). So high a ratio of details to achievements is simply enervating. The reader ends up in the position of the schoolchild who said, after completing a library report on penguins, that she now knew more about them than she ever wanted to. The authors characteristically over-record particulars without bringing significant achievements into focus. A reader cannot tell whether this is a second-order problem about communicating real achievements to an audience, or a first-order problem about not experiencing significant achievements in the first place, due to a relatively passive mode of experience. Students do mention working hard and facing up to obstacles and limitations, typically in writing, math, or seminar participation, but they rarely spell out these central conflicts and achievements, filling their text instead with book and activity lists which give the impression that they have worked only at the 'micro' level, like ants or bees, without focusing conscious attention on the shape of the whole.

But if circumstantiality is still a problem in average-rated evaluations, nevertheless it is a higher level of circumstantiality than in lower-rated ones. It is no longer about friction cars and geometric tiles, but is usually about authors and texts. Average students very frequently list books which "interested" them, often presenting the reader with a sort of guessing game as to what the student found compelling. A concrete example of this kind of excessive concreteness is Marx's *German Ideology*, Melville's short story "Benito Cereno," and Hawthorne's *Scarlet Letter*. A student enjoyed these three books most from a quarter's reading list, but gave the reader no indication of what shape of triangle she had drawn to connect them, or whether she saw them as connected at all. But at least one can say this is high-level circumstantiality, allowing the reader to conclude that Evergreen students read books.

The worst cases of focusing on particulars and omitting overall perspectives and objectives are those of seniors who claim that college has prepared them well for something, though the reader cannot figure out what. In a senior evaluation, a student begins and ends with aspirations about work and graduate study; but the statements sound inappropriately vague and are repeated more like incantations than plans: "Most importantly this program has helped prepare me for any future work or graduate school I plan to take on"; "Most importantly this program has prepared me for future study or work in any area that I become involved." The record indicates that the student had taken four successive coordinated studies programs, none of them advanced, ending in

the senior year with an entry-level or sophomore program. A reader of his self-evaluations cannot guess what sort of work the student contemplates, nor what field of graduate study he might be prepared to undertake. A vague unreality clouds the assumption that graduate schools await his application. Quite a few of the average evaluations leave the reader puzzled as to what disciplinary or interdisciplinary inquiries the student was pursuing, and why.

In general, the exhaustive enumeration of particulars seems to express what Freire (1970) calls the "banking" metaphor for education. If exhaustive enumeration is not just an evasion of the evaluative task, then it depends on some rudimentary theory of knowledge and learning, a theory which is *satisfied* by enumerations. Reflection suggests that the student is employing an epistemology in which 'more' is better and of higher value, and in which knowledge is information that accumulates quantitatively like capital in the bank. Thus in writing, the more one has said, the more one has demonstrated hard work and personal achievement. This is the same line of reasoning - usually implicit and latent, rather than conscious - that leads students to write library research papers that repeat all the information they could find, without developing a thesis.

Evaluations written in this manner, long and almost punitively boring, force the reader to skim or skip whole paragraphs, wondering if a grading system doesn't do a good enough job. Overall, such evaluations reveal inability to sort the important from the trivial and to organize memories and reflections around some criteria of significance. They are powerful evidence that *good writing wants a thesis*. As they stand, many such documents seem fundamentally dysfunctional, assuming their purpose is to help the student evaluate her achievements or to enable transcript readers to assess her talents.

Exhaustive enumeration without organizing principles occurs frequently in evaluations by students whose cognitive development is average, whose readings were extensive, and whose prose style is competent at the 'micro' level. One wonders if these students have been asked and trained in their programs to write academic essays with a thesis. Have they basically failed to master the idea of a thesis in any of their writing, or have they only secondarily failed to transfer this notion from academic essays to other forms of writing in which it would be useful? One cannot help but attribute some of the difficulty to lack of training as writers, in mapping the fallout of particular points into an organized topography.<sup>6</sup> Though nothing in this general account proves that writing instruction would reduce the tendency toward exhaustive enumeration of particulars without attention to organizing themes, the account at least suggests it would be worth a try.

---

<sup>6</sup>A reader familiar with Gilligan's distinction (1982) between stereotypic masculine 'linear' reasoning and (perhaps equally stereotypic) feminine 'network' cognition might suspect that some self-evaluations seemed inadequately organized to the raters because they were looking for the linear patterns and might have been ignoring the network kind. But close reading did not reveal networks of association underlying the default strategies.

**Organization: Two 'Default' Strategies**

The statement that the average evaluations usually do not have enough structure and organization to cue the reader and provide a sense of direction does not mean they are disorganized. They tend to be organized according to two 'default' strategies, akin to the values or processes a computer will use if not advised to do otherwise. The authors have indeed employed categories to divide and organize experience, but they are not using substantive or ideational categories, such as program themes. They tend to ask themselves into what activities and time-periods the program was divided. The average evaluation writer, to organize her exhaustive enumeration of particulars, typically falls back on the two structures which require the least conceptual effort: either an account of the program activities which comprise the weekly schedule (typically seminars, lectures, labs or workshops, and papers), or a sequential narration of activities quarter by quarter. Sometimes both methods are combined. In quarter-by-quarter evaluations, the time units are often used for organizing units even when the reader can see no thematic difference between them. Some of these are no doubt just quarterly, informal (or 'in-house') evaluations strung together, to minimize effort in writing the overall evaluation at the end of the year.

These two default strategies - organization of the text in terms of program activities and academic quarters - are simple enough that they do not need illustration. In interpreting them, the most important point is that they defeat other modes of organization. The default strategies defeat any articulation of program themes or contents which impacted the author, and defeat any real assessment of personal accomplishment. In evaluations using these strategies, one reads about whole universes of seminars attended and galaxies of papers written, without learning what ideas developed through such discussions, or what the papers were about. At some point, description of the activities of 'learning how to learn' defeats an exposition of actual learnings.

On the one hand, perhaps the activities and the quarters *were* the primary units of experience, and thematic contents just did not register: seminars, writing workshops, and lectures, or Fall, Winter, and Spring, were experientially 'more real' than the thematic content of the program as a whole, e.g., of Shakespeare or environmental studies. This is probably not what faculty members want to hear, but it certainly might be what students are saying. Unoriginal and formalized as such descriptions seem, they can, after all, be read as true and complete descriptive statements. Their message is that the student *is* describing all that she accomplished. Going through the routine activities *was* the extent of her achievement. She was a follower of the program schedule and did the things asked of her, but was not really engaged in inquiry and therefore was not deeply affected by the intellectual possibilities with which she was presented. It was perhaps all she could do, perhaps because of a job or a busy personal life, to keep up, to get some or most of the books read, to make it to seminars, and to fulfill the requirement for papers. Perhaps she really did not focus her attention consciously enough, or with enough feeling, on substantive themes to remember them well. If this were the case, nothing is 'wrong' with self-evaluations which describe only the program activities, quarter by quarter. They are thorough and accurate as they stand. The student really does not remember being moved by Shakespeare or having made discoveries about the environment, because such personal achievements *did not occur*.

On the other hand, the two convenient default strategies might have become bureaucratized conventions, equivalent to office forms or word-processing macros

which require only that one fill in the blanks. Perhaps the intellectual experiences were real enough, but hard to remember or recount, and so a laziness has taken over when it comes time to sort and evaluate them.

Such issues cannot be decided on the basis of available evidence. All that can be concluded clearly is that the results of these default strategies stand in the way of more serious modes of presentation.

In particular, these default strategies clarify or accentuate the distance between student and faculty experience. Whoever uses the activities of the week and the academic calendar to describe personal achievement is not in control, lacks an overview, and has not 'decentered' from immediate experience sufficiently to understand how or why the faculty put the program together. Viewed entirely from within, a program is like a dot puzzle drawn by adults, for children: the child follows instructions and connects the dots so that a picture emerges, without the child actually having created anything. Average evaluations tend to present academic experience in such a way, as if the student followed the steps but did not comprehend the design principles of 'the big picture' - did not perceive or address the disciplinary or interdisciplinary themes which presumably inspired the faculty team. Most faculty would be disappointed to read a batch of evaluations with an eye on the program themes which mattered to them. Average students, particularly those who use the default strategies, simply do not mention, except in rare breakthroughs, the major themes which the faculty probably thought they were presenting.

Together, these tendencies - first, to substitute exhaustive enumeration for evaluation, and second, to use the default strategies of organization - probably explain, on at least a rudimentary level, why composition and communication ratings improve so little in student self-evaluations. Composition included voice and structure, as well as standard English grammar; and communication centered on linking the particular *to the general*. The quantitative results showed how composition, which improves just a little, and communication, which improves not at all, tend to arrive at a plateau early in the student's academic career. The writing habits which have just been discussed are the symptoms, and at a low level of analysis are the causes, of this flat, non-developmental pattern.

These observations reinforce the case for writing instruction beyond Core programs. *This is where the average student is stuck:* in exhaustive enumeration, and in ways of organizing a text which serve better as labor-saving devices than as ways of communicating ideas. Students are stuck at this point in their writing, or at least stuck in their ability to transfer their writing skills to 'real world' tasks, at the level they tend to attain at the end of the freshman year. Two possibilities are that (a) these students did not learn to write a paper structured around substantive ideas (i.e., around a thesis); or that (b) if they did learn to do so when writing an essay, they did not learn to transfer this skill from the academic essay to 'real world' expository writing. Either evaluating one's accomplishments was not a situation perceived as requiring this sort of organized writing, or organized writing had not yet become a habit which could be transferred.

Writing instruction aiming at organizing ideas around a thesis would be helpful precisely because the problems in question - making lists of things, organized in obvious but not particularly meaningful categories - tend to become habits. Later, many

examples will be presented of breakthroughs from these problematic patterns; but the breakthroughs do not occur in all cases and, even when they do occur, usually do not reorient the basic pattern of subsequent writing. The evidence indicates that 'natural' stimuli do not consistently propel students to abandon these habits, nor does more practice drive such habits away. Rather, they are the kind of habits which, without intervention, tend to be repeated. Such habits can be applied to any writing task, and may even constitute the essence of bureaucratic writing, so that the student who has learned these habits is a 'finished product' equipped for a lifetime - if keeping track of lots of things in obvious but not very meaningful categories is what life is cut out to be.

But if a writing style and a quality of communication which bespeak authentic inquiry are goals of liberal education, then average students should receive help and stimulation to move beyond habitual constraints such as have been described. Two steps seem to be in order: first, to be sure that all students can write essays with theses or organizing principles; and second, to help them realize that focused inquiry, as opposed to enumeration and routine categorization, is not just an academic formalism, but is a real-world requirement for work of high quality.

### ***Procedural Knowing***

Just as average-rated evaluations have a typical writing style and organization, so do they have a typical "epistemology" or "way of knowing," that which Belenky et al. (1986) characterize as "separate," "procedural" knowing.<sup>7</sup> Procedural knowing is any mode that is dominated by a methodology. The two main procedural modes, according to Belenky et al., are "separate" and "connected" knowing. Separate knowing is the more linear and positivistic form, emphasizing objectivism and methodology; connected knowing is a more interpersonal, network-like way of 'getting to know' authors and ideas as if they were persons. Both appear in Evergreen self-evaluations, as indicated by the informal content analysis in the previous chapter. In particular, connected knowing appeared in relation to faculty, to authors, and in the codings called "Learning from peers" and "Seminar as process." But the great preponderance of codings fell in the category of "Procedural knowing." Instances of procedural knowing are much more often of the "separate" than the "connected" kind. This may be a surprise, since ideologically Evergreen claims to be the home of connected knowing in seminars and in faculty-student relationships, and of interdisciplinary rather than disciplinary emphases. But in self-evaluations, the mode of knowing which students employ and discuss by far the most extensively is separate, procedural knowing in the form of skills, techniques, and methodologies.

In this way of knowing, students make claims to know, or they look for signs that they have learned, on the basis of use of the methods of different disciplines, broadly conceived (i.e., "seminaring" and "writing," not just chemistry, history, sociology, etc., are conceived as skills or disciplines). The central activity in this mode of knowing is problem-solving. Problem-solving procedures take precedence over inquiry into relationships among concepts, methods, or systems. One has confidence that one knows when one has solved the problem at hand by applying the skill or method which provides right or good answers. The method or skill is a guarantor, invested with a kind of certainty.

---

<sup>7</sup>In the Perry scheme, this mode is closest to Position 3 (Appendix III).



On the one hand, this is a relatively advanced mode of knowing which seems to depend on college education: Belenky et al. found that it did not tend to arise spontaneously, without formal post-secondary education; and Moore (1982-86) reports that it is the dominant way of knowing among upper-division college students. But on the other hand, it does not reach out to include the cognitive complexity of more consciously relativistic thought. In procedural knowing, the different ways of knowing stand in relation to one another like the tiles in a "mosaic" of separate ways of knowing (Nelson, 1988, 1989). Or skills, methods, and procedures are like hopscotch squares: you jump into the right one for the problem at hand, and avoid crossing any boundaries. The knower is still naive about relations among what appear to be separate and discrete ways of knowing. Problems about how different methodologies and disciplines relate to one another have not yet come to the fore. Such awareness tends to occur later in the trajectory of cognitive development, when students encounter more complex relativistic thought and undertake "constructed knowing" (which takes uncertainty and relativism for granted and then appreciates or constructs contextually relativistic models and theories).<sup>8</sup>

Almost all Evergreen self-evaluations discuss some modes of procedural knowing, ranging from relatively naive efforts to learn "how to write" and "how to seminar," to thorough mastery of academic and professional methods and procedures.

The most typical statement about procedural learning is that "I worked on seminar-ing" or "worked on my writing." By an overwhelming margin, these are the skills most frequently mentioned. If students devote so much attention to seminar-ing and writing, either they are what mattered most, or they are what one can be discussed most easily, without arousing anxiety. But a difficulty arises when a focus on writing and seminar skills is achieved at the cost of reducing opportunities for saying *what* was written and talked about. For example, during the two quarters when a student studied utopias, one hears how she learned to write, and how much she participated in discussions; but one never learns what she thinks utopias are, what functions they serve, what problems they pose, or what her own utopia looks like. In many such instances, discussions of writing and seminar-ing have supplanted, rather than augmented, statements of what content was mastered. Learning-how-to-learn seems to have won too decisive a victory. The authors can even be said to have *reified* the methods and procedures of learning-how-to-learn, coming to treat writing and seminar-ing as ends-in-themselves, in much the same way that writing reports and holding meetings become ends-in-themselves of bureaucracies. It is as if Evergreen students, deprived of traditional majors, were covertly choosing to major in the two disciplines called writing and seminar-ing. Perhaps too much anxiety focuses on writing and seminar-ing, without enough clarity from faculty about how to allay it; or perhaps Evergreen or its evalua-

---

<sup>8</sup>See Appendix III. In the Perry and Belenky models, and also that of Robert Kegan (1982), more complex modes of knowing do not abolish those already in place, but relativize them so that they can be used when most appropriate, and in combination. In other words, a student who can formulate models and theories can still employ the different problem-solving methods of separate, procedural knowing, but she can also think of them as overlapping and as more or less appropriate in different contexts. She no longer regards the results of correct procedures as objectively certain. Rather, the relation of disciplines, theories, and methods is understood to be complex, relative, and interactive.

tion process actually promotes this reification of writing and seminarizing at the expense of a focus on content.

In some instances, however, the discussion of writing skills becomes much more specific and sophisticated, as in this example from an average-rated evaluation which contains a clear thesis statement and overall structure:

My writing skills when I entered the program were acceptable, but rough. In the process of the year I managed to first recognize and later write papers which were considered college-level work. In the beginning I had a tendency toward weak thesis statements and poorly organized papers. By spring quarter I was able ... to write a well organized ten-page paper on the philosophical belief of the [X] Church. This positive experience has made me a more confident writer.

This statement communicates successfully by connecting the general and the particular: the reader learns what the writing problems were, how they were remedied, and what was written about. It is a real evaluation, giving reasons why a particular accomplishment achieved standards appropriate for college-level work.

This is, in other words, an illustration of procedural knowing when it is especially well done. In many other fields besides writing and seminarizing, Evergreen students whose evaluations were rated average are able to offer well-developed examples of their mastery of methods and procedures:

I learned ... the importance of close reading when dealing with well-written literature. For example, a close reading of Robert Frost's poem, "The Road Less Travelled," exposed its true meaning, and it is not to take the road less travelled.

Some examples are esoteric, but indicate that the author is delving deeply into a particular discipline:

One of the main points ... stressed by my instructor and the ideas which I found most interesting, and at the same time difficult, was the idea of structured programming. This concept encompassed the use of logically arranged subroutines, meaningful variable names, thorough documentation (REM statements and tail remarks), and carefully planned algorithms, as well as the use of logical loops in place of iterative loops and GOTO's.

Statements such as this one may require professional or disciplinary readers, but this is fair enough: faculty members, personnel officers, and graduate admissions committees probably speak the appropriate dialects.

In these impressive instances, the nature of procedural knowing has not changed, but the level of procedural or methodological complexity certainly has. These students are no longer working on the seminarizing and writing skills of 'going to school,' but are making contact with the skills and the modes of discourse of academic and vocational communities beyond Evergreen. They have not yet mastered interdisciplinary discourse, nor have they attained what some would regard as the highest goals of liberal arts and sciences education; but their personal achievement does

them credit. Important disciplinary components of learning and knowing are clearly in place and procedural knowing - the modal "epistemology" of American college graduates - has been mastered. When such students say they have attained skills useful in work or graduate study, the reader now knows what fields they are contemplating and in what ways they are prepared.

### ***Something to Say: Breakthroughs into Ideation***

In evaluations written by students who rank at the top of the middle decile, and also in those of upper-division students throughout the decile, seedlings of intellectual interest begin to break through the surface. If the overall tone of average-rated evaluations is like a bureaucratic accountability report, and if the documents are therefore not very engrossing for the reader, nevertheless things start to happen. Nearly every packet of four years' evaluations contains at least one passage which one can call a 'breakthrough into ideation' - a statement of substance indicating that the student achieved some overview or inspiration, some rapport or connection with an engrossing idea, some creative insight or achievement in arts or scholarship that stands out from the ordinary and is really a personal achievement. The evaluations would not have been rated overall as average if these breakthroughs were the organizing principles of the evaluations, or if the insight were sustained in a series of texts; such lasting breakthroughs would have elevated the overall rating to another decile. But the point is that they do appear in average evaluations and can be taken to represent the authors' highest personal achievements in the academic sphere (or *vita contempliva*). Even within sets of documents describing a generally bureaucratic mode of school-attendance, one finds evidence that Evergreen has provided the space for experiencing and recording distinctive personal accomplishments.

These accomplishments are the same kinds as appear in higher-rated evaluations, where they are featured more prominently and extensively; and so they need only brief characterization in this section.

1. **Organizing ideas:** Moving toward the high end of the average-rated evaluations, the most noticeable difference is the appearance of one or more central points which stand out from the description of routine activities. These do not always appear at the beginning of the evaluation, to guide the reader; and they are often not explained adequately in the text, so that they remain mere assertions which the reader cannot fully appreciate; and they are rarely used to organize more than one paragraph of the text. But the writer has made some personal statement or decision about what mattered most. Such evaluations are 'better' in the functional sense of serving to communicate what the author thought was an achievement. To begin with, they are interesting enough to be read rather than skipped over. In addition, they are functional by at least four contextually relevant criteria: they reveal the student as an active evaluator of her own educational experience; they help the faculty member to learn what mattered to the student; and they help external readers of the transcript to gain information which is more complete than a list of activities; and finally, they make more sense to readers when they appear under the heading of "The Student's Own Evaluation of Personal Achievement."

2. **Statements of purpose:** In the fifth decile, one encounters statements of purpose and orientation - a way that the mind anticipates, rather than just follows, experience:

I enrolled in [an environmental studies program] because I felt that it would directly address my concerns and questions about the state of our planet and its possible futures.

The author, already as a freshman, is a purposive agent able to remember or reconstruct aims that led down a curricular pathway. The self is an intentional, planning, choosing agent, in contrast to a passive recipient of experience.

3. **Overviews:** When the following passage surfaced at the beginning of an evaluation, it told the reader just what had been missing in the many low- and average-rated evaluations that had been read before it:

[Program X] set out to "unite" the hard sciences with the humanities: to align the rationale of Modern Science with the human morals and values represented in the humanities, which play such an influential part in human thinking and action. The need for this was well conveyed in the course.

Here the writer has finally stepped back and has posed a crucial issue - perhaps *the* crucial issue - not of which experiences to mention serially, but of *what the program was about*. This allows the reader to put the author's achievements in context and perspective: does the reader understand some significant connections or disparities between the two cultures of sciences and humanities? Granted, the evaluation does not answer this question, but the essential act of posing it has occurred.

4. **Self-awareness:** In some of the average-rated evaluations and even as early, as in the following example, as the sophomore year, the self is subject to observation (in writing), and the author becomes aware of and able to communicate changing states of the self:

For me, this quarter has been one of some change in attitude and perception of myself. Through the weeks I've noticed a really devoted individual taking my place, working vigorously toward personal goals and standards. I've also found that my strict attitude ... caused some people to get mad at me, so I really don't know if this change is for the good or bad.

Such self-awareness can go beyond puzzlement into awareness of larger themes in one's studies:

This quarter's work has had a great impact on me as an artist. An awakening of awareness has occurred. For the first time I see an ordering of apparent consistency in my work and perceptions. Ironically, this order seems to be based on the notion of conflict.... The concept I take away from these studies, is that in art as well as in life, *conflict* is the dynamic mover of most events. Conflict is interesting. It is realistic because it is something we are easily able to believe in. It can be applied to all art.

Though the inference is speculative, perhaps one can think of this kind of overview as a product of the evaluation process itself, or of the "culture of evaluation" at Evergreen (Thompson, 1989).

**5. Before-and-after views, or learning from experience:** A special instance of overview combined with self-awareness is the before-and-after self-portrait, which allows one to say how one changed and what one learned from specific experiences. The following example presents the process in rich detail. A history student followed his faculty member's advice, constructed a time-line of relevant events, and then made a mistake she does not plan to repeat:

The next logical step, then, would have been to identify a specific group of people, ... [and] discuss how they interacted with the trends.... But somehow I missed this step and instead tried numerous general approaches.

By the time I came to terms with the fact that my approach ... was ... generating unsubstantiated conclusions, I was so enmeshed in generalities that it was difficult to extricate myself and begin again. But I did begin again and although the final paper was short, it is the germ of something, that with time and effort, could be developed into a long, thought-provoking paper....

The process of working on this project was exciting and informative, and I am confident that if I undertake a similar project in the future that I will not make the same mistakes.

At the procedural and disciplinary level, the author has learned something important about historical method and has gained a genuine sense of the historian's vocation. She now understands that applying other people's general approaches without confronting a specific body of data results only in ideological interpretations which buttress one's preconceptions; and that the only way to 'listen to the data' and learn something new is to identify a population and confront real evidence. But at an even more important level, this involves a *structural* change in the way the student thinks. She has taken a big step in 'cognitive space,' so that she can take a retrospective overview and pass a judgment - a real evaluation - on what can now be seen, though it could not have been seen before. The paragraphs contain two positions or perspectives - one, in which the naive self bumbled down the wrong path; and another, from which a later, more sophisticated self looks back at what happened and sees where the naive self had wandered into the dark wood of error. This two-dimensional before-and-after perspective did not appear in the lowest decile of evaluations. Those were one-dimensional testimonials; this is more complex exposition of the evidence of learning.

**6. Commitments:** Breakthroughs into ideation in a few instances take the special form of commitment to what one discovers one thinks. Seminars, when they work well, call for saying what one thinks, and by this process students often discover there *is* something they think, even when others do not. This kind of breakthrough into ideation justifies the seminar process. It occurs with heightened anxiety in art critiques, where the student, showing her work, has to hang up "a bit of my soul" on the wall:

Another new experience for me was having my [art work] critiqued. I will probably never forget the feeling when I hung my watercolor self-portraits;

I felt I had hung a bit of my soul up for everyone to comment on. The comments ... were not all positive, but they were constructive and I learned firsthand what it means to continue something, despite what others say. I knew then that if people had thrown tomatoes at my work, I would still continue to draw and create.

This student's breakthrough into ideation is a real breakthrough into commitment, a discovery of what she is moved to go on doing whether others like it or not. This is an especially impressive achievement because it turns up in the senior evaluation of a student who had consistently had difficulty completing assignments. Before, assignments had come from the outside; but finally a motivation has sprung from within.

**7. Interdisciplinary perspectives:** Average evaluations also express intermittent awareness of the nature of coordinated or interdisciplinary studies:

Being in a coordinated study program, involving a wide variety of disciplines, was new to me and I enjoyed this variety and the fact that the central ideas which we tackled could be and were approached from so many angles. Previously my interests were mainly in the [lab] sciences and so this program gave me exposure to many different viewpoints, ... particularly ... environmental science and sociobiology.

In average-rated evaluations, only a few references such as this indicate that the student was aware of involvement in interdisciplinary studies. The evaluation does not stick to this theme, but at least it begins with it. It also contains a characteristic twist: the student, though aware of being in a coordinated studies program, still seems to be saying that it was not the relationship between viewpoints - the interdisciplinary perspective - but the plurality and distinctness of points of view - the disciplinary perspectives - that impressed her. If this is what she means, then it typifies the way the world divides up into a mosaic of discrete parts, from the viewpoint of separate procedural knowing.

**8. Theoretical thinking:** Finally, the breakthroughs into ideation which appear in average evaluations contain instances of sound theoretical thinking. But the word "instances" needs emphasis: if theoretical thinking could be sustained in writing, then the evaluations would have rated much higher. The following examples are not developed beyond the quotations offered here. They are not necessarily even the central point of the evaluation from which they are taken. They are not sustained within the same evaluation, nor matched in level of complexity by material in previous or subsequent evaluations. But they could not have been written without an emergent capacity for theoretical thinking - for "contextually relativistic thought," in Perry's terms (1970), or for "constructed knowing," in Belenky et al.'s (1986). Three examples from average-rated evaluations illustrate theoretical thinking emergent in three forms: questioning, analysis, and application of theory to practice.

a. *Theoretical questioning:* A student of revolutions noticed that all Socialist revolutions, though egalitarian in intent, have been led by an elite and have contained a power structure.

This power structure has been a major problem with so-called socialist revolutions. I cannot conclude that such a course is invalid, for revolutions seem

always to require some sort of catalyst. I ask, for the people to directly gain control, must it be such a group, ... or can a catalyst be generated by the people as a whole? This leads me to the ultimate question of, can people live and act together as a country, without a small governing state? Is a ruling class a requirement for forming any kind of coherent social structure?

This student, still only a freshman, is talking on paper about substantive ideas, consequential ones. She says she can come to no conclusions, as yet; but the important point is that she is doing what none of the evaluations in the first decile could do: she is making herself the arbiter, the one who must figure out what makes sense and what doesn't, regarding a question of theoretical interest and practical social significance.<sup>9</sup>

b. *Theoretical analysis:* Most procedural knowing - indeed, all examples of it in the low and average-rated evaluations, except for the following one - treated methods as freestanding 'givens,' something given by nature or authority: one just uses them, like divine or natural laws, without understanding how methods follow from theories and connect them to data. In one instance, however, an otherwise average-rated social science student<sup>10</sup> learned to see how theories, methods, and data are connected, how theoretical frameworks generate hypotheses which are in turn supported by data:

[A faculty member's] advice on critical analysis brought focus and direction to my reading. By distinguishing the theoretical framework from the supporting empirical data, ... I was able to establish a clearer perspective from which to analyze the books. I now consciously approach texts in search of their theoretical basis and how well such hypotheses are supported.

c. *Theory and practice:* Finally, one student working in a cinema production group was able to take a recently formulated theory and invent a new way of putting it into practice:

Briefly, Psychological Montage is a newly formed concept that takes the five Eisensteinian levels of [film] editing one step further; to a *sixth* level. It is an attempt to more closely integrate the audience with the on-screen drama, action, etc. The methods used to achieve this closer association are: Overall length of film, repetitive actions/events, other editing or cinematic devices which are known to bridge the distance between character and spectator.

Further, ... we used a technique we would later call "frustrated camera" movements, which consist of a camera that seems to search for its shot, and sometimes unsatisfied, will change angles abruptly....

The conclusion of this evaluation effectively concludes this section, as a model of 'breakthrough' learning:

---

<sup>9</sup>She has, in fact, rediscovered what the early political sociologist Roberto Michels (1915) called "the iron law of oligarchy."

<sup>10</sup>This is one of the few cases of substantial disparity among ratings, with a high cognitive development rating (Position 4) occurring in conjunction with composition and communication ratings which were low enough to bring the student down to the middle decile.

To speak more broadly, I've never taken a personal theory such as this, and put it to the kind of test that we have done here. It has been a very unique learning experience to actually work physically with some very abstract concepts.

### ***Teaching Writing With Substance***

To sum up these impressive breakthroughs which reach into matters of substance, one can say that they probably exemplify what most faculty hope most students are getting from their programs, and they probably also exemplify moments which were especially gratifying to the students themselves. These are the high spots in the everyday work of students who are neither dazzling writers nor consistently complex thinkers, and thus cannot really be called intellectuals. Mostly they have put in their time 'going to school,' writing about their experience in the form of bureaucratic reporting; but finally they have given evidence of what they have understood. When this occurs, the examples are impressive. Though the examples did not serve as organizing themes of the evaluations in which they appeared, each indicates that more was going on than bureaucratic or time-serving business-as-usual.

If faculty have leverage in the process of writing and thinking, they should apply it to stimulate change of this kind. *If such change does not occur, the evidence is that students remain on a plateau of little accomplishment in writing and thinking.* One can go even further and say that this is the point where education inevitably shades off into formation of character. It is fashionable to say that everyone - including quite young people - is entitled to whatever character she or he just *has*. Admitting that education involves character-formation may even seem reprehensible, on the assumption that characters are formed especially in the matrices of gender, class, and ethnicity; nobody should have this uniqueness - or group conformity - modified from the outside. These are very confused points in modern and post-modern pedagogy. But as Aristotle noted, education really is a matter of *forming habits of mind*. If a student forms the habit of mind in college of avoiding evaluative thinking by substituting what amount to bureaucratic forms, this is presumably the habit of intellectual performance that will be carried into life outside college. If, on the other hand, she forms the habit of making 'breakthroughs into ideation,' formulating ideas and dealing with issues of substance, this too is a habit that applies beyond college. Much is at stake in terms of the values of liberal arts and sciences education. The student who stays on the plateau of the default strategies, who lists and enumerates without really identifying personal achievement, is either avoiding or having difficulty with *deliberative thinking* - with what Aristotle called "practical wisdom" (*Eth.* vi.5), the sort of thinking that helps us figure out what is to be done. If one does not become habituated to this kind of thinking - if it does not become part of one's habits and character - then such questions are simply left to others, and the individual is that much less autonomous, that much less a 'real' individual, and that much more "banal" (Arendt, 1977) in what we have come to call life-style. In this context, the difference between the student as bureaucrat and the student (the *same* student) as someone who makes breakthroughs is central to the educational enterprise. One can see the difference in the self-evaluations under discussion. Difficulty with deliberative thinking is visible at the composition level, as absence of a thesis in the organization and content of writing; at the communication level, as difficulty in deriving something general to say from the particular; and at the



level of cognitive complexity, as evasion of the central task of deciding what is important, or of exercising judgment. The breakthroughs, wherever they appear - in seminars, papers, evaluations, conferences, wherever - are the *prima materia*, the best of the raw material that a faculty member is given to work with.

This point of view leads first of all to a recommendation about teaching writing. In the previous section, which concluded that many students need help with writing after the freshman year, the argument was that their problem is more than technical, and involves *finding something to say*. The examples just considered are relevant instances of students having done so. They are a series of substantive statements generated by the average students themselves, illustrating what they *do* say when moved. The list of eight categories - admittedly not exhaustive of having something to say, but certainly illustrative of it - can be reviewed as the sort of things on the tips of students' tongues, the next statements that they might make if moved or prompted. They deserve review by anybody teaching writing, as models of what to expect and encourage. The Soviet psychologist Lev Vygotsky (see Bruner, 1986) has argued that intellectual performances too advanced for a student to appreciate them do not provide instruction, encouragement, or modeling, but are just a kind of static or are even intimidating. In contrast, Vygotsky formulated the concept of a "Zone of Proximal Development" (ZPD), representing the area just one flight up, so to speak, which the student could reach if teachers helped by providing the ladders or scaffolds to reach it. The eight kinds of examples here illustrate what lies in the average student's Zone of Proximal Development.

But secondly, these examples were not used to organize the texts in which they appeared. Here is an especially important place for writing instruction: teaching students how to use their own emergent ideas as central or structuring principles for their writing. The 'thesis thesis' states that good writing needs a thesis, or some organizing principles, or at least a point of view to cue the reader and organize the text. This abstract statement now becomes concrete and is linked to the particular: the kinds of statements which have amounted to breakthroughs into ideation for average students are the kind that can be developed into theses or organizing principles in their expository writing.

Finally, writing instruction must cultivate sustained and transferable skills. This requires habituation over a period of time. When the potentialities which have been located have been developed into organizing principles in writing, then the habit of writing this way must be *sustained*. One breakthrough, as the average evaluations indicate, does not form a habit. Students have to get used to the process of recognizing and amplifying 'something to say.' In a circular process, one has to begin with some rudimentary idea and must then proceed to work it out on paper. *Finally, a habit of thinking on paper is what is missing, and would have to be cultivated for overall quality of writing to improve.* This is probably why so little improvement appeared in the quantitative studies of the composition and communication scales: if a student does not establish the habit of thinking on paper as a freshman, then writing does not substantially improve.

### ***A Case for 'Summative' Self-Evaluation***

Reading the average-rated evaluations suggests another recommendation, not about teaching writing and thinking, but about the evaluation process itself. The middle decile of self-evaluations, though grammatical, readable, informative, and sometimes exciting when breakthroughs appear, is above all boring. One can read a pretty large batch of them without becoming involved or being moved to take notes. There is nothing 'wrong' with them; there is just not that much which, in the long run, matters. Everything that every writer mentions happened, and perhaps mattered, to a real person; and if one looks with a microscopic eye for signs of development, one can almost always find such signs. But this does not contradict the general impression that these evaluations tend to bore their readers.<sup>11</sup> In general, in average evaluations, the detail greatly exceeds or even overwhelms the interesting points put forth in them. The reader's eye seems to have come into focus on too small a world, where one sees in too much detail the slow developmental motion of average young people leading relatively immature and uneventful intellectual lives. The effect is like time-lapse photography that moves too slowly for one to see the bud open.

This arises first from mistakes in the assumptions the writers make about the audience's level of interest. If the audience is one's own memory bank or the faculty member sitting across the room in the evaluation conference, then a case can be made for considerable detail. But if the audience is readers at a later time, in personnel or graduate admissions offices, then the level of detail in average students' transcripts is self-defeating. Second, the authors are too concerned with presenting exhaustive lists of activities, whereas the reader is probably looking instead for *changes* and *developments*. Third, the two 'default strategies' seem to inhibit focusing on intellectual development. Fourth, the writer often incorrectly assumes that a report on what "interested me" will interest the reader, which is often not the case, because the writer's interests are too undeveloped or immature. Finally, many developmental trajectories render their own starting-points irrelevant: the freshman and sophomore years would for many students best be left only briefly recorded.

All of these problems might be overcome or minimized if Evergreen instituted summary evaluations written at graduation - or 'summative' evaluations, as these are called in educational jargon. These could replace the longer and less functional documents. To preserve the pedagogically important practice of end-of-program self-evaluation, a summative evaluation might be accepted only if it superseded end-of-program evaluations already on file.

Such evaluations are required, and are currently being assessed in terms of cognitive development, at Fairhaven College of Western Washington University (M.

---

<sup>11</sup>It is of course possible that the boring quality attributed here to average evaluations follows instead from research fatigue. The researcher had to read stacks of evaluations all the way through, whereas an admissions officer or prospective employer reads only one. And the researcher has to read them serially from the freshman to the senior year, whereas the reader-in-the-field receives the transcript with the senior year on top, and needs to persist only as long as the material engages interest. But even with this qualification, many attributes of these evaluations imply reader boredom. Circumstantiality and lack of an overview, for example, probably do not make a good impression on anyone.

Eaton, personal communication). Fairhaven requires narrative self-evaluations at several points before graduation, but the self-evaluation included in the transcript is one written during the senior year. Initial research findings indicate that the summative evaluations tend consistently to receive higher MID or Perry ratings - in the context of the present study, higher ratings for cognitive complexity. The overall gain and final results are almost exactly the same at Fairhaven as at Evergreen, but Fairhaven's 'senior surge' is even larger because of the procedure of summative self-evaluation.<sup>12</sup> Summative evaluation thus seems to serve the pedagogic value of inducing the sort of overview or decentration that was lacking in average Evergreen evaluations.

The main policy question about summative evaluations would probably be whether they would strengthen or weaken the current procedure for end-of-program evaluations. Counsel from Fairhaven College might be obtained on this point. Fairhaven seems to have kept both procedures working with synergy. The most that might happen is that Evergreen summative evaluations might display the same elevation in ratings which the Fairhaven researchers found. The least that would happen is that students who still wrote only bureaucratic accountability reports would end up with only one long document in which to bore the reader.

### **The Top Decile and Engagement of the Self**

What had appeared as blips of ideation in the otherwise routine reports of average students become dominant themes in the highest-rated evaluations. In some instances they become the organizing principles of all or part of the evaluation. These are evaluations *per se*, displaying a more thoughtful sense of what it means to evaluate one's own personal achievements. Though procedural knowing remains prominent, these evaluations also contain what Belenky et al. (1986) call "connected" and "constructed" knowing, extending to a capacity to understand and to construct one's own models and theories. These evaluations are full of substantive statements displaying what was actually learned in the way of new ideas. And they are chronicles of change, developmental in nature: the authors are able to observe change in the self and communicate a sense of change to a reading audience. Finally, these evaluations are the work of students who are *intellectually committed*, those who study not just to master procedures and fulfill expectations but to experience their own engagement, as knowers, with the known.

### ***The Theme of Change in a Freshman Evaluation***

The best way to experience the dimension of change in the higher-rated evaluations is to take a close look at the developmental process described in just one evaluation. The example is the freshman evaluation from the set rated highest overall.

The structure of this evaluation is entirely different from the average ones. The purpose has shifted from chronicling or recording activities to explaining the main

---

<sup>12</sup>Dividing regular program evaluations into three time groupings of Early, Middle, and Late (because they are not always written at the end of academic years), Fairhaven's mean MID ratings were 3.34, Early; 3.42, Middle; and 3.65, Late ( $n = 126$ ). But the mean summative rating was higher, at 4.16.

dimensions of change. It begins by going back to the beginning, to initial expectations, in order to show how the expectations themselves have changed:

I had hoped that this course [in modern social science and intellectual history] would provide me with basic academic skills, and fill in the gaps in my understanding of modern Western culture.... My own reading ... prior to college had taught me to sense patterns in the chronology of history, and I expected this class to refine thinking that I'd already done.

The writer initially expected more of the same - in the banking metaphor for education (Freire, 1970), incremental deposits, with interest, in existing accounts. Implicitly, his initial idea was that Western cultural history conformed to the Idea of Progress, and thus justified an optimistic view of the present and future. Change came when he realized that neither the world's economy nor humankind's interaction with the environment functioned benignly, and that their trends might point as directly towards catastrophe as towards progress. His whole world-view shifted as a result of taking his freshman program readings to heart. The change came when he read about Enlightenment ideas of progress through science, but realized that the human condition is no better than before. Even Marx's idea of revolutionary change seemed utopian and false:

Karl Marx proposed that Capitalism was the inevitable precursor to an ideal and equal worldwide society of cooperation. The change would come, he explained, when capitalism had oppressed the worker so much that he would rise up with his fellows and overthrow the bosses. After reading *Germinal*, I wasn't so sure this was possible. The bosses controlled wages, and people needed to eat. I felt disillusioned and concerned for the fate of the human race.

This disillusionment about the human condition, originating in reflection on Zola's novel, deepened when he began to imagine runaway rational-bureaucratic social organization and environmental catastrophe.

Modernization has brought with it a tremendous faith in science and rationality.... [But] ultimate technological destruction in the form of atomic holocaust or being suffocated by our own effluents is unacceptable to me. So is a homogeneous, rational society of totally adapted human beings. This class has made me very pessimistic about the probable fate of modern Man but [has] given me the skills ... to seek alternatives.

Instead of the refinement of analytic skills and the filling of knowledge-gaps which the author had anticipated, this reorientation is complete. The change, depressing as it may be, marks reorientation in intellectual life, in social expectations, and in commitments of the self to action. It is not merely analytic, or all in the head, but is a change in what the author *cares* about and in the ways he thinks of *acting*. It is an emotional or affective as well as a cognitive reorientation. The two have gone hand in hand, with intellect and feelings united. The author is not only more knowledgeable but more worried about the world in which he lives. The change also involves a capacity to *dialogue* with the authors and to create a conversation of ideas. The dialogue is not just about facts, but about values. The formerly positive values of science and rationality, if not actually reversing, at least come into question as he observes how science and rationality lead to their own excess. One 'master narrative'

has replaced another one, so that the Idea of Progress, with its implied optimism, has yielded to a sense of doom. The author accepts this new darkness on grounds that his new worldview is less self-deceptive; and he still takes knowledge, even gloomy knowledge, to be a condition of empowerment in seeking valued alternatives and trying, at least, to resist destructive ones.

The basic structure of the evaluation is a before-and-after contrast of two modes of thinking. It is an exercise in 'metathinking' or thinking about thinking, one which delineates the kind of thought that went before and the kind which has superseded it. The subject of the evaluation is thus not just the activities of the program or the books read, but the *change in the student's whole way of addressing the human condition*.

Furthermore, this intellectual and personal reorientation is *aligned with the structure of the program*: the student has made his own discoveries, but they are the result of considerations which the faculty had meant to raise in devising the syllabus. The student is, then, thinking along with the faculty, appreciating the program themes they had embodied in the program. He is perhaps not yet - while still a freshman! - creatively free to devise his own worldview, for in a sense, he has traced out a big connect-the-dots puzzle which the faculty had embedded in the reading list. But the faculty probably do not themselves have the answers to the questions he is raising, and so in this sense, he is thinking at the same level as they.

This is the kind of thought and the kind of change one encounters in a top-rated evaluation. It is an integrated structure of ideas based on the relationship between the program and the self. It derives from academic experience the concepts which describe the author's own year of extraordinary growth. It is as fundamentally different from the average-rated evaluations as those were from the low-rated ones.

### ***Many Forms of Lucidity and Insight***

Overall, this kind of intellectual engagement of the self distinguishes the top-rated decile of evaluations from the others. No one way predominates as a way of involving the self. Again, no *Pilgrim's Progress* narrative arranges the following points as a series of stations through which all student journeys should pass. The discussion does not try to say that all themes, or even more than one, appear in all of the evaluations; only that views of change involving the self take such forms as these.

1. **An active orientation:** In tone, these evaluations speak actively, rather than passively or bureaucratically, and sometimes virtually crackle with energy. In addition to completing a demanding senior group contract in psychology, the second-highest-rated student reports in the senior year that

Along with the anxiety of graduation upon me, I chose to also work 20 hours a week as a resident advisor/student manager, teach Red Cross swim classes, swim on the college team, participate in a six-month internship ... as a family co-therapist, and take part in a three-month internship ... [in] outdoor education. Now, let's look at first quarter.

Actively oriented evaluations like this one contrast to the bureaucratic orientation by welcoming novelty. Low- and average-rated evaluations did not contain

celebrations of the unexpected. The authors essentially wanted more of what they had encountered before - a mere accrual, according to the banking model, of more of the same intellectual capital. If anything, they tend to recoil against the new and unexpected, perceived as sources of trouble and inadequacy. In contrast, high-rated evaluations are full of statements of initial expectations, followed by statements like "the course has been more than I ever anticipated, introducing me to new concepts...." A note of excitement, of appreciation of intellectual stimulation and novelty, separates these evaluations from the others, indicating willingness to participate without serious reservation in the life of the mind. The welcoming attitude toward stimulation and novelty extended even to considerable risk-taking:

The highlight of the whole program was the final performances.... I thoroughly violated my comfort zone and did some things for the first time in my life. In the first place, I spoke to an audience and maintained my confidence and composure all the while. I also performed a solo song and sung lyrics I wrote. The theme of my performance was "risk-taking in art," and every aspect of that performance violated my comfort zone, giving me a totally new self-image ... as a performer. The audience's positive response and standing ovation was wonderful and overwhelming.

No statement comparable to this appears in the low-rated or average deciles.

2. **'Decentration':** Self-awareness, in these evaluations, takes on much greater complexity as the authors increase their capacity to stand back from the present, to compare and contrast before-and-after images, and to chart the changes in the self. Decentration is not just an analytical stance involving objectification; rather, it can be an intuitive process, in the Jungian sense of seeing patterns of implications and possibilities (Jung, 1921). This student claims to be aware now of what had previously been an unconscious preoccupation:

... Through my development the past year ran the luminous strand of religion. It showed in nearly everything I did, from an essay on sacrifice to a tendency, an aesthetic desire, for "religious gestures" in movement I choreographed. I was not brought up in any particular religion, and I subscribe to none now, so the development was purely natural. Recently I read that individuals in our culture have been "orphaned by a civilization whose mystery is exhausted." Perhaps the author had been looking for this "mystery" in the wrong places. I felt that here, through the medium of theater, I was constantly trying to express some sort of mystery, "known only through vague hints, but essentially unknown...."

Through self-observation, some students are also able to identify processes of their own behavior which they believe (rightly or wrongly) that they can now shape or bring under self-control:

A lack of confidence, even an unwillingness to formulate strong personal viewpoints on what I read and heard, was something I considered a flaw of my educational process. I imposed standards I had to work on.

3. **Goal-setting:** In such instances, students are setting personal goals and priorities. Authors of low- and average-rated evaluations tended to regard all program

requirements as equally important benchmarks established by purely external authority. All instances of falling short of expectations thus tend to be logged as failures in authority's eyes and as indicators of inadequacy of the self, calling for guilty confessions or excuses. But higher-rated evaluations more often introduce personal criteria authored by the self, in the form of goals and priorities. In this vein, the student can then say, for example, "In the process of exploring these goals I have completed most of the program requirements, though I learned to set priorities." The student means 'I didn't do all the work,' but this can be said without implications of failure, because the work was not some sort of external bureaucratic requirement. It was instead a set of intentions authored by the self, with some priorities more important to the author than others.

4. **Evaluation *per se*:** Decentration and goal-setting enhance the possibility for evaluation as such, as the student becomes adept at comparing and contrasting different times, concepts, and contexts. The task is to note differences (personal achievements) and to evaluate, i.e., place values on them. The higher-rated authors tend to understand clearly that the self-evaluation document *is* an evaluation. Their evaluative statements are far more complex and communicative than the 'pre-evaluative' low- and average-rated statements of likes, dislikes, and interests. In some instances, evaluation is a complex process of relating the self to the themes of the program:

There are two, not always distinct, levels of evaluation which I will attempt to reconcile. These two levels are my personal development - which includes academic development as ... part of a greater, ongoing whole - and secondly, my "participation" - a salient mode of learning in ... academic activities but also in the whole of my ongoing life.... As the program approached culmination, ... I could detect linkages between course-determined criteria for "development" and personally-determined criteria....

In a program on human development, this student saw that the program content was relevant to the self and useful in the process of evaluation; and so program themes and concepts actually come into play within the evaluation.

5. **Critical thinking:** The low- and average-rated evaluations displayed mostly subjective thinking (what I like, what interests me) and procedural or how-to-do-it thinking. The above-average ones introduce critical thinking that focuses on reasonableness of assumptions and propositions, and relative merit of differing points of view. At the level of reasonableness, specific arguments might be rejected - e.g., "I found this argument difficult to swallow," because "it seemed judgmental and assumptive." And different arguments can be compared and contrasted to one another - a procedure which appeared in none of the low- or average-rated evaluations:

The [performing arts] program showed me that the performance process, from creation to product, can be a medium for growth for the performers involved. In other performances I have been a part of, emphasis was placed on the final product, with little importance given to the process itself. Although this is the typical professional attitude, it does not take into consideration the needs of the performers as creative beings, or concentrate on expanding their potentials. [This program], by encouraging students to experience the process of ... creation, ... sparked personal transformations of attitude and outlook in myself and my colleagues....

This student, working with a particular experience, considers it in relation to the consensual position that product is more important than process. And he does not end up in a doctrinaire advocacy of the process orientation. Rather, he concludes in a spirit of *inquiry*: "What sort of transformations can be effected through performance art in the audience and in myself, the performer?" This sort of "logic of inquiry" is a sure sign that critical thinking is functioning well (see Cederblom & Paulsen, 1991, re: the contrast between "debater's logic" and the "logic of inquiry").

**6. Capacity for generalization:** In several ways, the higher-rated evaluations reveal a capacity for extended generalization, or for better communication by virtue of linking the particular to the more general. Whereas lower- and middle-ranked evaluations tend to follow the program moment by moment without abstracting the general from the particular, the higher-ranked ones reveal considerable capacity for generalization and abstraction, usually well anchored in program themes and texts. For example, a freshman in an interdisciplinary communications program concluded that

... Technology is advancing, ... [but] human expression is being overshadowed, even dominated, by technical artifacts which separate man from his own and others' creative impulse because of their convenience value. It's almost as though there were an invisible shackle on individual and collective creative processes put there by man's serving a cultural commitment to a mechanistic ... view of the "world."

**7. Larger contexts:** In several ways, the high-rated evaluations testify to changes in the relationship of the knower to the known. The simplest is just to see issues in a larger context. An example is provided by a drama student:

Perhaps the most striking thing about my participation in [a performing arts program] was a change in orientation I went through from fall to spring quarter. My interest slowly shifted from acting to theater as a whole, in other words the standpoint of the director. As I was exposed to various media of expression my focus expanded. This does not mean I intend to stop acting. It means I am thinking about theater in a comprehensive way.

This example illustrates again that decentration and reorientation of the knower in relation to the known need not imply objectification or intellectual distancing. The student remains personally involved, but now appreciates a larger view of a discipline and a broader range of action for the self.

**8. Awareness of program themes:** A specific instance of greater capacity for generalization, and also of awareness of broader contexts, is recognition of a program's overall interdisciplinary themes. For example, "Our reading list exposed me to a contrast which was to become a recurrent theme in this course: the difference between the interests of the individual and what had been determined as best for all of society." In a sense, students who apprehend such overall themes are the only ones functionally enrolled in interdisciplinary coordinated studies, pursuing and appreciating the themes which the faculty have used to structure the program. Other students appreciate the ideology of interdisciplinary studies as 'a good thing,' likeable and interesting, or they accept such studies passively, as what was offered; and no doubt all benefit in the long run from exposure to connections among disciplines, even if they do not reflect exten-



sively on such connections. But it is especially those students who commit to paper their reflections on interdisciplinary themes who are attuned to the goals expressed in the College's mission statement, and who are studying what the faculty purport to be teaching.

9. **Connected knowing:** Belenky et al. (1986) have suggested that some students - especially, they believe, women - will not be as attracted to separate procedural knowing and will work more from the perspective of networks and connections. This too, they argue, is a procedure: the subjects or objects of knowing are being treated as if they were people, and one 'gets to know' them by procedures modeled on interpersonal relations. In the present body of data, separate procedural knowing was the rule and connected knowing the exception. In particular, very little connected knowing appeared in the lower and average groups of evaluations. But connected knowing appears in the higher-rated evaluations, first, in connection with peers in seminars; second, in relations with faculty; and third, in relations with authors and texts.

a. *Connecting with peers:* As content analysis revealed, the higher-rated evaluations are more likely to contain evidence of learning from one's peers, and awareness of seminar dynamics as observable and alterable processes. The authors of high-rated evaluations are much less concerned with finding their own voice, because they already 'own' and bespeak an identity. They are free in this sense to be more concerned with their seminar as a whole. They become conscious of seminar as a process, and think of 'good' and 'bad' seminars less from the standpoint of personal expression and more in terms of communicative processes involving everyone:

In seminar, I developed my ability to question non-threateningly, a valuable tool for unearthing unspoken assumptions. Also, it is hard work to involve everyone, but it paid off in diverse perspectives, and the feeling of unity that comes when everyone's involved. Another important group skill I discovered is to take a clear stance on a topic. A well-defined statement is easy to understand and counter. A nebulous one tends to stall discussions, to obscure the focus.

The student's perspective in this passage is outside of the self and its immediate needs. The value of seminars is holistic or contextual, and depends on full expression of a diversity of opinions. Diversity and unity are values, but the statement goes beyond this. It goes on to suggest a Socratic sifting of opinions, figuring out which are tenable when the clearly defined ones are sorted out from the nebulous.

b. *Connecting with faculty:* In the higher-rated evaluations, faculty are no longer what Perry (1970) calls Authorities with a Capital A. They become lower-case authorities, people with whom one talks if not exactly as an equal, then at least as a member of the same species who is only more experienced in certain kinds of inquiry. Faculty members are seen as participants in the same process of inquiry as one pursues oneself. They know what they know not because they wrote the tablets of knowledge or even because the tablets belong to them, but because they have been translating or interpreting for a longer time. The tablets belong equally to the student, and the faculty member has become more a mentor or guide. In some cases, authors of high-rated evaluations have worked very closely with faculty, either in small group contracts

or on research in which the faculty was interested.<sup>13</sup> In these circumstances, faculty are less authorities than peers and co-workers at a professional level of inquiry, e.g., "At the start of this year I chose to continue study with Professor [X], helping him to devise a critical apparatus for the academic investigation of traditional verbal arts."

c. *Connecting with authors and texts:* The most notable instances of connected knowing in these documents take the form of dialogue with the authors of program texts. Texts are not authoritative, and are not just repositories of information; they are *people talking*, or dialogues in which the authors address the student, and he or she 'talks back.' The student who noticed the program theme of conflicts of interest between individual and society went on to say that

The first view was John Locke's - that society is a mere tool to protect the individual and his property. Jean-Jacques Rousseau exposed us to an alternative view that Man enters into a social contract and must then bow to the will of the majority.

Such dialogic processes are complex indeed. The student has accepted a particular kind of *construct* or *fiction* about program texts. The texts are not just there to read sequentially for historical information, but are *dialogically juxtaposed*, as if the authors were speakers in a conversation, and the program reading list and its readers were included in a larger conversation-of-the-whole. This viewpoint embodies what Belenky et al. (1986) mean by "connected" and "constructed" knowing. Instead of just listing the books he liked or found interesting, the student hears a conversation. The writer sees Locke and Rousseau as offering theories and arguments and as taking positions. He learns *from* rather than *about* them. The authors are active agents with speaking voices, as in the phrase "Rousseau exposed us...." And the implications of this dialogic process extend far beyond the context of these particular readings. One can hear and understand almost everything - science, social theory and data, artistic movements - in this manner, as dialogue within communities of speakers. This student has become accustomed to a way of reading, thinking, and writing that changes the relation of knower and known, and the whole structure of intellectual activity, into dialogue.<sup>14</sup>

---

<sup>13</sup>Among this small group of graduates, close associations with the faculty occurred in the humanities and sciences. The evaluations conveyed the impression that arts faculty tend to sponsor advanced projects in which the students work more as a collaborative production group, and less as apprentices to the faculty. No close collaborations with social science faculty appeared in this particular group of evaluations.

<sup>14</sup>Faculty characteristically assume that this idea of dialogic juxtaposition will come easy, but students frequently find it difficult and even complain that they have been reading too many texts that were 'wrong.' For example, in a recent program on "Meaning, Learning, and Power," Don Finkel, Burt Guttman and the present writer included in the syllabus five "educational ideals" drawn from classical Greece, Zen Buddhism, the Enlightenment, the American democratic tradition, and contemporary feminism. Many students, probably a considerable majority, assumed that these ideals were related only sequentially and narratively, not dialectically, and that each one was supposed to supersede and erase the previous one, in a history of conflicts with a final victor (ethnic and feminist). The ideas that these positions might also address or 'talk to' one another, and that together they might comprise a dialogue in which all of the voices have something valuable to say, did not come at all easily.

**10. Constructed knowing (theories and models):** In a previous illustration, a student of human development described two perspectives, personal and academic, which could be brought together within one evaluation. As the student makes these statements and actually performs these operations, the reader is witnessing the kind of complex thought which Perry (1970) calls "contextual relativism" and which Belenky et al. (1986) call "constructed knowing." The evaluation is much more than a report on achievements already accomplished. Like other especially thoughtful evaluations, it is an achievement in its own right - a creative essay and a dynamic example of the complex thinking processes which the student has learned to employ. Thus evaluation has itself become an act of "constructed knowing."

Perry's concept of "contextual relativism" (1970) and Belenky et al.'s concept of "constructed knowing" (1986) describe an "epistemology" in which the student acts as a theorist or builder of models. The student has realized there are no absolutes, no one right procedure nor one right assertion that is absolutely or objectively true. But she has got over the naive relativism which results in a solipsistic kind of "subjective knowledge" (Belenky et al., 1986, ch. 4). Instead, she realizes that while nothing is absolutely true, one can construct theories and models which are more or less representative of situation and context. Moreover, one can examine such theories and models and invoke quantitative and qualitative evidence to decide which representation is, relatively speaking, more appropriate to the context at hand.

From this perspective, a student is no longer forced to opt for one method or another by the hop-scotch method of problem-solving, merely jumping from one method to another without being able to relate them. Now she can reason about the appropriateness of different methods (or combinations) and modes of representation to different contexts, as in this passage:

I am not one to explain or intensely interpret the "meaning" of the work I've done. I shuffle images into patterns which seem to create a coherent performance, whole in itself. Images have logic peculiar to themselves, and we can perceive their "artistic validity" (or non-validity) kinesthetically and intuitively. To explain or interpret image or dance theater with language patterns used to discuss literature I believe may lead to confusion over the nature of the work.

With a capacity for contextually relativistic thinking comes a stronger motive for critical reasoning. Premises are no longer arbitrary, but have to be spelled out so that the reader can judge their merit. Theory suddenly make sense: it is not just the story told by Authorities, but is the best statement one can make about why one's premises are appropriate to their subject matter and context. Statements of complex and well-reasoned theoretical foundations thus begin to appear in evaluations:

I began with the premise that performing traditions (Homeric epics, British ballads, jazz, nursery rhymes, etc.) have similarities in function and mode of transmission which are distinct from those in literature (or classical music), because literature is preserved by standardized notation systems for ready storage and retrieval, whereas knowledge which is not so stored must rely on certain conventions of repetition and formulas in order to be remembered and passed on.

While contextually relativistic or constructed knowing prompts critical thinking, it is not, however, exclusively 'cool,' analytic, or rationalistic. The whole point is that when nothing is objectively or absolutely true, one has to achieve some clarity about the "experiential" relation of the knower to the known (Lakoff & Johnson, 1980; Lakoff, 1987). In this spirit, what Belenky et al. (1986) would no doubt recognize as constructed knowing enters into this performing arts student's creative project:

Fear and the unconscious mind were the two major concepts I dealt with in putting together my staged, final performance.... Using dreams - nightmares - as a vehicle I attempted to illustrate how, if we deal with them rationally, our unconscious insights can be consciously grasped and ultimately contribute to, or even become, the foundations of our creativity.

**11. Theory and practice:** In the discussion of breakthroughs in average evaluations, one example demonstrated that a film student was able to translate theory into practice. In higher-rated evaluations, this skill is more widespread and occurs at a much higher level of generalization, as in this example:

A fundamental motivation for me this year continued to be the perception that the dominant industrial societies are extremely out of balance with the natural processes of Earth that make life possible. More than ever, my life's passion is in catalyzing the economic, political, and cultural turnaround that must happen if we're to minimize catastrophes and reclaim an appreciative relationship to Earth and to each other. It can be shattering to look objectively at the difficulties we face, but, having made the commitment to a hopeful future, I've approached these problems as challenges.

This student has taken an overview that allows her to conceive broad, even global, problems and objectives. She has related theory to practice by considering the difficulty of acting against the odds. Less realistic students want to retain a simplistic, world-saving optimism, and often feel deprived of efficacy, or even feel angry at knowledge itself, when their studies begin to teach them that the world is not ready to change in the way they wish. In contrast, this student has thought over the predicaments and obligations of deciding what to do when the ethically best action has the odds stacked against it. She is willing to make a commitment to action even if success is unlikely. This is her challenge. Her statement is the best example in the cases at hand of the committed and practical idealism for which Greeners have earned respect.

**12. Creative process, or control with freedom:** Students who can handle contextual relativism or constructed knowing have abandoned certainty and its rigidities, and thus are the first students in this study who are truly free to experiment creatively with ideas. Methods and procedures have a new meaning, in that they can be open-ended: they don't guarantee incontrovertible results, but they do provide one with the discipline that leads to the cutting edge or to the state of the art. Some students are able to work with open-ended procedures, with control exercised over a larger process of emergent results. A theater student (one who was quoted before) exemplifies this sort of freedom. He describes directing a project in which actors were assigned characters in a partially formulated production and were then encouraged to find the "central gestures" of their characters. They then interacted in an impromptu performance of those gestures. "The participants ... allowed themselves to descend in

a sea of emotions and memory and remain there as those experiences formed themselves into gestures delineating a story." The result was an animated "collective explosion of emotions," which the student director then dramatized to produce the final production. As inventive and fruitful as such a process sounds (and as conventionally *avant garde*), the point is that such ideas simply do not appear in lower- and average-rated evaluations. Their authors struggled to grasp linear procedures, but in no instance described a multi-staged, open-ended process, one in which the inquirer or creator exercises control not over anticipated outcomes but over emergent patterns that manifest themselves through a kind of chaos theory. The writer lucidly describes his approach to this kind of freedom and control, and the firm position he took to maintain and direct it:

I didn't realize how successful I'd be in this undertaking or what I was getting myself into. [Actors] became so involved in the personal emotional aspects of the experience that [they] became nauseous during rehearsals ... [or were] depressed.... I'm a novice director, this was a difficult situation for me. I felt I had to maintain a strong level of faith as others looked to me for support and assurances. Since I had from the beginning worked from a John Cage inspired style of non-interference in a process, I refused to let myself or anyone else retreat from the water and short circuit the process by composing a nice, safe theatrical piece. I believe that by allowing ourselves to experience fear and anxiety we gained ... a dimension of depth in our work.

This blend of freedom and control was just not a possibility for most students. It exemplifies a mature kind of creativity that lets a writer or composer or scientist experiment as freely as possible, without imposing an interpretation on outcomes until interesting results have emerged. It can be described legitimately as a creative process.

**13. Autonomy within the curriculum:** The previous themes have emphasized cognitive development, whereas the next ones move out of cognition narrowly conceived, into the areas where cognitive development and personality development evolve together. (These are the areas Perry designated as Position 6 and above; see Appendix IIIB).

Average- and high-rated evaluations differ considerably in their approach to the curriculum. For average students, the curriculum is a given, and the temporal perspective is annual, or quarter-by-quarter. In contrast, the students whose evaluations receive higher ratings do not choose programs as if choosing channels, simply in terms of what interests them next. Instead, a language of curricular planning and structured choice enters their evaluations, as in this example:

My agenda for next year is already in formulation - in sequence. In a program based on visual communication, I hope to explore other modes of intelligence and expression besides the written and the verbal analytical modes. From my past and recent experiences I believe in the possibility of eventually using all the forms of experience and knowledge in a therapy built on creative expression....

In such sets of evaluations, the reader can almost always see the pattern and understand the reasons for moving from one program or contract to another. In contrast, in many lower-and average-rated evaluations, the reader simply cannot grasp or follow a pattern. The higher-rated students are not just *in* programs and do not just *take* them,

as relatively passive recipients; instead, they *use* programs as vehicles for accomplishing broader personal and academic objectives which are stated consciously and which persist through time, with a sense of purpose and planning. In general, the quality of thinking involved in this sort of active use of the curriculum expresses a life-stance. It contains the assumption that one lives in the world as more than an experiential onlooker or "received knower," and that one at least attempts to exert some conscious control over the sequence of stimuli one receives.

**14. Authoritativeness:** The evaluations which received high overall ratings introduce different notions of authority. As mentioned before, teachers are no longer alien others, or Authorities with a capital "A," and have become knowledgeable people with whom one dialogues in order to appraise their positions. But what increases most impressively is the high-rated students' *own* sense of authority. They gradually assume a mantle of authoritativeness and speak with the confidence appropriate to ('lower case') authority. This is usually not an overt act, but just something the student slips into. It usually appears in the form of language identifying the self with a group of skilled intellectuals or professionals. The word "studying" disappears: the student stops referring to herself as someone studying history or chemistry, or some interdisciplinary field like environmental studies, and starts talking about herself as a person who *does* what historians, chemists, or environmentalists do. For example, "Because the ability to focus on the music as it occurs is the musician's number one priority, I have disciplined myself to concentrate better during every performance and practice." The 'writing ego' is no longer a music student, but a musician - a person whose actual achievements allow her to include herself in the professional category.

**15. Graduating oneself:** Graduation tends to appear in these evaluations as a meaningful event, a rite of passage, not just a turning of the last page of the academic calendar. In other evaluations, the reader frequently cannot tell from the text that the student is graduating. They contain no 'final act,' no drama, culmination, or stock-taking. In contrast, the high-rated senior evaluations rarely treat the final year as more business-as-usual. They tend to have a stock-taking, cumulative and summative quality, so that at least part of the final evaluation turns into an assessment of change in college as a whole.

Some seniors have gone out of their way to bring the undergraduate years to a close with a peak or culminating experience. In a few cases, most often in the arts, a student creates a personal rite of passage or 'graduates herself' with a special ritual.<sup>15</sup> High-rated performing arts evaluations often described directing a whole film or production in the senior year, and studio artists described presenting individual or group shows as culminating experiences. A professional musician, for example, ended his studies with a performance in which he was composer, bandleader, instrumentalist, and M.C. He graduates saying that "This is the first time in my musical career that I have conceived and actualized my own concept of a musical performance," and "I can't think of a better way to end my studies at TESC." Several students were consciously taking on, in projects or internships, roles for which they hoped to be qualified upon

---

<sup>15</sup>Though some students write senior theses, no instance of a formal thesis surfaced in this group of evaluations.

graduation - for example, psychological counselors interning in mental health agencies or in the College's Counseling Center.<sup>16</sup>

**16. Life after Evergreen:** Finally, where average students seem to graduate and then encounter whatever happens next, the writers of high-rated evaluations more often chart deliberate trajectories from undergraduate experience into life after Evergreen. Besides a vocationally purposive internship, this might involve a plan for graduate study, or even more extensive life-planning. In the high-rated evaluations, while the student's competence may extend beyond one discipline, the reader can still tell at what sort of graduate program the student is aiming. In terms of broader life-planning, the most energetic of students - the one who studied psychotherapy, held a job, taught and competed in swimming, and undertook two internships - ends his undergraduate career

enthusiastic to continue ... doing research and ... to continually redefine my psychotherapeutic outlook.... I am on the road to finding a niche in the world of psychology and I'm excited to continue learning.... In the near future I plan to be working with emotionally disturbed children in an outdoor education program,... then ... to travel ... and live and teach [abroad] for a year, then enroll in graduate school and receive my masters in education and my Ph.D. in educational psychology.

This is the extreme case of life-planning, and this life may even seem over-planned. Since things change, such comprehensive planning might be considered a luxury or a delusion. But it expresses a sense of agency and a stance towards action which, whatever the future holds, provide the author with the sort of "dream" which helps young adults to open the doorways of possibility (Levinson, 1978).

### ***The Engaged Self***

With a little bit of irony, the contents of the highest-rated evaluations have been presented in the form of a long list. The directions taken in these evaluations are plural. Overall, they take shape as a garden of forking paths. Precisely because they are less uniform and are individually more interesting, they resist assignment to a few categories. They do, however, have a common quality, and that is *engagement of the self in the process of education*. For these students, education does not just happen to one, but is something one does for oneself. These students have broken the bureaucratic educational mold which had tried to contain them from K through 12. They have hatched through the shell of formal education as active, purposive individuals. Ironically, it is not the subjective mode of knowing - or the person who wants to be known through her likes, dislikes, and interests - that conveys a sense of personhood to the reader, because the subjective preferences and interests of youth are not yet sufficiently differentiated to be very interesting. The most interesting students in the sample are those who make themselves known through the content or substance of education, conceived especially as change in the self and its capacities. The reader cannot help but follow their progress with lively interest. For education is, after all - and as Socrates

---

<sup>16</sup>None of the students in this group mentioned staying in her internship as a paying job, though this is the idealized or most desired instance of 'graduating oneself' by transforming successful undergraduate experience into 'life after Evergreen.'

would have it - a matter of shaping and presenting one's character (Finkel & Arney, 1992).

The general argument of this paper is, then, complete: First of all, many students after the freshman year seem to need help with their writing. When their writing is weak, it is weakest in want of a thesis, or of organizing principles. But having a thesis is not a mechanical or technical writing problem. It is a matter of having something to say. And having something to say, it now seems, is a matter of taking oneself and one's relations to learning and knowing very seriously. Helping students to formulate a thesis, to find something to say, to engage the self: broad demands indeed are involved in teaching writing and thinking.

### **Endnote on Decline and High Gain**

As an endnote, mention should be made of some special cases outside the three deciles which have been studied closely. One group of cases involved decline in overall rating, and others involved especially high gain.<sup>17</sup> The former do not prove to be distressing, and the latter give the reader a good deal of hope about how far students can travel in their writing and thinking at Evergreen.

#### ***Decline***

Fortunately the instances of overall decline are not alarming, because they tended to be strong, intelligent writers who understandably became increasingly involved in practical and career-oriented interests, thus tending to write in later evaluations about more circumstantial achievements and skills. These were not internal dropouts or 'remittance men' or other kinds of students in name only who had escaped the demands of intellectual life.<sup>18</sup> Two cases of overall decline illustrate this point. One was an older woman who changed her career interests, launched herself in the arts, and then had to figure out practical ways to earn a living for herself and her daughter through application of her newfound talents. The earlier evaluations describing breakthroughs in the arts were rated higher than the later ones about studying economic survival tactics. Another example was a student who intended from matriculation to enter medical school, but wanted to begin with a broad liberal arts and sciences background. Her early evaluations were thoughtful essays on far-ranging intellectual experiences, while the later ones were less exciting chronicles of her fulfillment of medical school prerequisites. In such instances, pragmatics necessarily triumphed over the modes of thought and expression appreciated by the rating systems. But declining

---

<sup>17</sup>A close reading of cases of disparity between high ratings on one scale and low ratings on another revealed nothing of interest. The reason is that the disparities did not turn out to be very large.

<sup>18</sup>Only one student whom one might regard as an Evergreen alternative version of a fraternity boy or sorority girl appeared among the indigenous graduates. This person seemed never to have got much of the program work done, so that a reader could wonder why so much credit was granted. But this was a single exception. While such students might enroll for a while, they do not seem to make it into the population of indigenous graduates.



ratings did not connote dropping out and did not necessarily indicate loss of intellectual energy or commitment.

### **High Gain**

A few cases are models for all the others, and it will be reassuring to end by looking closely at a few of these. Some students did what one might have hoped all students might do: they increased their overall rating by distances that required moving an average of a whole step forward on all three scales, e.g., from "Adequate but undistinguished" to "Good" on the composition scale, or from one whole Perry position to the next.<sup>19</sup>

Because the high-gain evaluation sets begin with different individual benchmarks, any effort to generalize is hazardous. In two or three out of eight instances of high gain (conceived as more than one standard deviation on the combined index), students received higher ratings without achieving what has been characterized as a 'breakthrough into ideation.' This could occur if the initial rating was extremely low or the writing by itself (i.e., composition and communication) became much cleaner, without much increase in cognitive complexity. One case may say something anecdotally about students who enroll for as little attendance as possible at the college, and take almost all individual contracts. This case reveals a prodigious amount of work - more than one paper over 50 pages, and one over 100. But the student seems only to have been a successful capitalist according to the banking metaphor for education. Much information must have been accumulated, but the evaluations do not describe ideas that are very interesting. It is possible to move ahead substantially in overall rating while remaining bounded by the horizons of procedural knowing.

In the other five instances, the students who achieve high gain also achieve major breakthroughs in terms of having something to say. These are the exceptions in which one can see what did not appear before - a kind of Pilgrim's progress. In general, the low, medium, and high-rated evaluations are three categories, not three steps in a progression through which students pass developmentally. But several of the students who achieved the highest overall gain have reproduced the qualities of the low, average, and high deciles in their own earlier, middle, and later evaluations. Three examples will convey an impression of what is accomplished starting from different baselines and improving substantially in writing and thinking.

1. *From skills to understanding:* The student who, in the whole population, was most preoccupied with procedural knowing and skill-building ranks third highest in overall gain because of a final breakthrough, as a senior, into substantive understanding of the political system of democracy. She was concerned almost to the point of obsession and compulsion with describing the skills she was acquiring, but had virtually nothing to say about the subject of all the programs she chose (modern European

---

<sup>19</sup>Because of the procedure for compiling the index, it is impossible to say whether or not gain in overall rating was significant. The reason is that when standard scores are computed, both the freshman and the senior means are zero. But if one simply averages the raw composition, communication, and cognitive complexity scores for freshmen and seniors and then compares them with a *t*-test, the senior mean is significantly higher, with  $p < .0005$ .

history). In the way that one might describe paintings as applications of brushstrokes, or piano playing in terms of fingering, she treats modern history as a neutral ground for skill development: time management, public speaking, expository writing, and construction of research papers. Then finally, in the senior year, comes understanding. The evaluation begins with a thesis statement, that "I have come to learn ... new ways of looking at history and an understanding of what democracy really means, at least to me." The body of the evaluation contains the following amplification of the thesis:

I have also come to have a more solid idea of what "democracy" might mean.... Prior to this class, I had always seen democracy as a governmental term, ... used as something that was opposite, and better, than communism and socialism.... We were all living in a democracy, and I never really thought about it beyond that. [But] in *The Populist Movement*, Goodwyn related the story of a revolt of a group of poor farmers, whose purpose was to change and improve their lives. This idea, of the people of the country being able to gather and work together to change their way of life, is what I think democracy is all about. Democracy is essentially about the people, not a term used by the government to show how much better we are than communists.

This thesis is announced near the beginning, so that the reader can keep it in mind as an orienting principle while following the details. The reader finally finds out *what the student learned* - the component that was so noticeably absent in her earlier evaluations, which had reified the processes of learning and had overlooked themes, ideas, and content. Perhaps the skill-building was necessary, but the payoff comes only when she finds something to say. Then the reader's attention snaps into focus, because she was able to *say something* about democracy as an abstraction generalized from the activities of particular people changing and improving their lives. She uses her insight to organize her whole statement of personal achievement, with the example of democracy providing evidence for a concluding statement that

I feel that this program has been a good one to finish my undergraduate college years. I feel that this school has taught me how to think more analytically, see more than one side of an issue, and how to listen to others and benefit from their insights and knowledge.

A reader of the full evaluation is bound to find this conclusion convincing.

2. *From high school to a vocation:* The second-highest instance of overall gain illustrates that achievement does not depend on a head start. In this instance, high gain depends on an initially low baseline. The freshman evaluation received the lowest ratings on composition and communication and was limited to descriptive statements about remedial work at the high-school level:

[Basic math] was a good course for me. It was more than just a review. I have a much better understanding of math now than I did before the course. I learned the importance of being able to solve algebraic equations and how to plot points on a graph. I can also do any problems dealing with fractions, decimals and percents.

In subsequent evaluations until the senior year, the texts become longer. Accomplishment is procedural and quantitative, not ideational (e.g., "I was able to

finish all the homework and do the final exam"). She summarizes programs by their component parts, e.g., seminar, math class, computer lab. The learning style typifies separate, procedural knowing, which testifies in this instance to considerable progress. She suffers difficulty expressing abstract ideas and calls herself a "descriptive" rather than an "analytic" writer. Apparently her analytic papers have been criticized for reading like book reports. She expects harder work, not more reflectiveness, to overcome her difficulties. By the junior year in a management program, she has overcome the inarticulateness of her freshman evaluation, and this is no small achievement. But her growth seems to have hit a ceiling, so that she seems to function (and to be preparing to function after graduation) as a competent managerial bureaucrat, with no perspective or viewpoint looking beyond application of the procedures she has been taught.

Finally as a senior, a change appears, albeit tentative and unsustained. She presents a point of substance, a conclusion she has reached herself:

Admitting that I am a descriptive writer, I feel I have made substantial gains in becoming an analytical writer. A paper on the auto industry was my best attempt at writing analytically. The paper focused on the auto industry's media hype and its attempted comeback. The argument I formulated was that this was a media blitz and the auto industry had not made the necessary sacrifices to become truly competitive.

This case exemplifies much-greater-than-average achievement which starts from the lowest benchmark. College has transformed deficient K-12 skills into competent bureaucratic-vocational skills. Some analytic capacity has emerged, in the form of ability to make one's own argument. How well the analytic skills develop in relation to the vocational ones depends on opportunities in post-collegiate life.

3. *From the 'default strategies' to multicultural awareness:* Finally, the student who registered the highest overall gain seems to have recapitulated in a developmental sequence the qualities of low-, average-, and high-rated evaluations. Initially, as a freshman, he organized his evaluation by type of program activity: seminar, papers, lectures, etc. The freshman evaluation listed books "I enjoyed," and describes them in terms too global to be informative: "These classics were interesting because of their content and what they said about society." This and later evaluations dwell on the fact that writing is difficult for him. In the sophomore year, however, he was able to state a thesis at the beginning of the evaluation, about skill development and improved writing. The organization of the evaluation changes, in that the discussion of activities is now preceded by a conceptually more thoughtful overview of what disciplines were addressed in the program. He has moved, in other words, well up into the level of the average ratings. In the junior year, he more successfully linked the particular to the general, and he can report that his academic work had some substance: one learns that he wrote comparative papers about feudalism and capitalism, and about early modern political economy. In the junior and senior years he set (but did not seem to achieve) the advanced goal of attaining integrated understanding of the themes of his interdisciplinary programs.

As a senior, his special accomplishment is a mini-essay, within his self-evaluation, about teaching English as a second language. The discussion effectively links the particular with the general, and theory with practice. He explains how his internship

taught him about cultural differences, particularly in concepts of time and in differing degrees of respect for traditional education. And he explains what he has learned about the plight of refugees in this locale. He has demonstrated, in other words, that he has triumphed over writing problems, that he can discuss issues of considerable depth, and that he can address these with personal insight. He has mastered real problems with composition, to the level of *presenting a thesis*; he can communicate, at the level of *having something to say*; and his essay on cultural differences bespeaks *engaging the self* in learning and knowing. In these respects he presents a model instance of what a young person can gain from college in terms of ability to master prose conventions, to communicate articulately, and to deal with increasingly complex ideas in writing.

These three cases are not necessarily the College's most impressive graduates, the ones who would emerge in a less equalitarian system as the Phi Beta Kappans or valedictorians; but they represent the College's most substantial success stories, the students who - one can say retrospectively - had the most to gain. The proportion of students who make this sort of progress is not very large, but no evidence suggests that other colleges and universities are accomplishing more than Evergreen. On the contrary, the separate study of cognitive development and the data cited from Fairhaven College indicate that Evergreen is ahead of most institutions and exactly at par with an 'alternative' peer institution. The three examples of high gain perhaps offer a little extra faith and hope to the whole faculty at Evergreen, who are involved in teaching writing and thinking.

### **WHO IS ADVANTAGED, WHO AT RISK?**

The previous section has described and interpreted, qualitatively and 'phenomenologically,' what Evergreen student writing and thinking look like, with a sample stratified into lowest, middle, and highest deciles. Now, with some knowledge of what writing and thinking look like at these three levels, one can pick up quantitative methods again and ask if any particular groups of students are advantaged in attaining high ratings according to this index, or are at risk for not progressing so far along the three lines of development that were measured.

The overall answer is that some group differences are significant, but they tend to be small. Statistics textbooks (or good ones, anyway) always discuss the difference between statistical significance, which refers only to the likelihood that the results might have appeared by chance, and commonsense significance or importance - whether the differences which are statistically significant are also important, from the standpoint of understanding and policy. This study reveals some statistically significant differences, but since they are not large, the reader has to determine how important they are. It does not reveal any differences concerning class and ethnicity. On the one hand, this is a good sign; one hopes that higher education has the effect of undoing class and ethnic disadvantages perpetuated in the rest of society. But on the other hand, the data were not robust. The measure of class or socio-economic status was not income, because that information was not available, but 'hereditary' versus first-generation college attendance. The criterion was whether or not one of the student's parents had attended any college. This proved to be too blunt an instrument to differentiate anything.<sup>1</sup> And the number of persons of color in the population of indigenous graduates was too small for statistical applications - only 12, and of several different ethnicities. Only the grossest of inequalities would have registered statistically under these circumstances, and it means little to say that the grossest inequalities were not revealed. The results were more interesting, however, in relation to other academic and demographic variables.

### **Factor Analysis: Preparation and Learning**

In looking for patterns among the different variables in this study, the place to start is with factor analysis, which reveals two underlying structures. One is what students brought with them to college in the way of preparation, and the other is what the index measured - what one might call the learning measure.

---

<sup>1</sup>Among the students' parents, 71% had attended some college - at least a course - and 29% had not. The imbalance in the size of the groups tended to defeat statistical tests, because only gross differences on other variables could register as significant. A more reasonable criterion would have been whether or not a parent had graduated from college, but this information was not available.

A first glance at all variables in the study (the three scales, academic background, demographics, and modes of study at Evergreen) fails on a measure of sampling accuracy (Kaiser-Meyer-Olkin measure of sampling adequacy = .41), but offers tantalizing clues. An 'achievement factor' seems to center on the three scales, which load positively with one another and negatively with age. A 'preparation factor' connects SAT and GPA with composition and communication - the more compulsive aspects of good work in high school. A possible 'specialization factor' loads group contracts (positive), coordinated studies (negative), and female gender. And a possible 'socio-economic factor' centers on parental education and suggests, albeit crudely, that perhaps the sons of higher-status families come to Evergreen with weaker high-school backgrounds. These are not significant findings, just clues to consider.

Looking at the three scales (using the average for all four years) and at demographic variables, and eliminating variables which do not load significantly on any factor,<sup>2</sup> the KMO measure of sampling adequacy rises acceptably to .73. The general pattern reduces to two factors (Table 3):

---

**Table 3:**  
**WRITING, THINKING, AND DEMOGRAPHICS:**  
**ROTATED FACTOR MATRIX**

	FACTOR 1: "INDEX"	FACTOR 2: "BACKGROUND"
Age at graduation	-.42	-.18
Gender	-.04	-.82
High-School GPA	.26	.68
Cognitive complexity	.86	-.16
Composition	.81	.31
Communication	.79	.44

---

<sup>2</sup>Ethnicity, presumably because of too few cases; parental education, because the criterion of having attended any college courses at all was set too low; and SAT, because of self-reporting, resulting in less data and in apparent underreporting of low scores.

One factor is organized by the combined index measures of composition, communication, and cognitive complexity, with loadings (relationships) for age and high-school GPA; and the other is dominated by gender and high-school achievement, with additional loadings for writing (composition and communication, but not thinking). This analysis indicates that one overall factor shaping the results in this study is what is learned in college and is measured by the index of writing and thinking; but the other is something students bring with them to college, a preparedness factor, and particularly a tendency of Evergreen women to have been higher achievers in high school. Communication, composition, and high-school GPA cross-load, i.e., are represented by fairly high numbers in both columns, indicating a kind of bridge between the two factors. In other words, the evidence tends to validate the reasonable assumption that early-formed writing habits form an important foundation for college achievement.

#### **Academic and Demographic Variables: GPA, Gender, Age, and Mode of Study**

These generalizations can be brought into tighter focus. The academic and demographic data gathered for this study can be used in conjunction with the combined index to look for students who are more likely to be high achievers in writing and thinking (as measured here) and those who are less likely to be. Four variables - gender, high-school GPA, age, and mode of study - are related at statistically significant levels to writing and thinking. The differences measured on these variables are all fairly small at the same time that they are statistically significant. They are summarized in Table 4 (page 80).

The data indicate that women, students who did well in high school, younger students, and students who take group contracts are more likely to be higher achievers in writing and thinking. Men in general, students who did not do so well in high school, older students, and students who did not take group contracts - by inference, those who took less advanced programs - are likely to rate lower in writing and thinking.

The essential qualification about the population applies here, that these are categories of the *indigenous graduates* - not of the student body as a whole, which contains many transfer students. These are the students who represent the College's maximum impact, but they do not represent all students.

The margins of difference are small enough and variability is high enough so that while one can say which *groups* within this population tend to rate higher and lower, nevertheless one cannot predict that any *individual* is distinctively at risk.

By constructing separate versions of the combined index for each year, freshman through senior, one can conclude that the group differences are generally greater at matriculation than at graduation, suggesting that Evergreen education has an equalitarian tendency. The differences tend to be ameliorated a little more, year by year, except for age differences, which do not decrease. With the exception of these age differences, the College tends to mitigate the small group disadvantages that existed initially.

---

**Table 4:**  
**VARIABLES RELATED TO COMBINED INDEX:**  
PREPARATION, GENDER, AGE, AND MODE OF STUDY

VARIABLE	TEST	RESULT	PROBABILITY
High-School GPA	Correl.	$r = .35$	$p = .007$
SAT	Correl.	$(r = .28)$	$(p = .07)$
Gender	T-test	Women $M = .24$ Men $M = -.17$	$p = .03$
Age at graduation	Correl.	$r = -.34$	$p = .002$
Coord. studies	Correl.	$r = -.25$	$p = .03$
Group contracts	Correl.	$r = .26$	$p = .02$

---

### ***High-School Preparation***

High-school grade point average (GPA), the primary measure of high-school preparation, indicated that students who were high achievers in high school tended to maintain a lead in college, albeit one that diminished through time. The correlation between high-school GPA and the combined index of writing and thinking was .35, a moderate positive correlation accounting for 12.4% of variation on the combined index ( $p = .007$ ). The secondary measure was Scholastic Aptitude Test score (or an older Washington equivalent test, with scores converted). Reporting of this score was voluntary, and so the number of cases was small (58) and reporting might have been inaccurate. The correlation was positive and small ( $r = .28$ ), but fell just short of statistical significance ( $p = .07$ ). On the one hand, such small and fragile correlations do not say that high-school achievement 'determines' achievement in writing and thinking in college; but on the other hand, they do show a connection which, in the case of GPA, is not likely to be due to chance.

As explained at the end of the section on the combined index, one can also construct a similar index by combining the three scales for each academic year, freshman through senior (though these measures are not protected as well as the average index against influences of faculty and peer editing). Following this procedure, the



small advantage of the high-school high achievers in writing and thinking persists through time, but diminishes year by year and loses its significance for upper-division students. The correlations between high-school GPA and a freshman combined index is .38 ( $p = .003$ ), .29, sophomores ( $p = .03$ ); .25, juniors ( $p = .055$ ); and .23, seniors ( $p = .09$ ). Again, the pattern indicates amelioration of group differences during the four years of college. Given the 'natural' equalization which already tends to occur, one cannot designate high-school underachievers as a disadvantaged group.

Looking more closely, one sees that the link was not equal for all elements of the index. The correlations for composition and communication, the aspects of the index bearing more on writing, were higher (composition,  $r = .41$ ,  $p < .0005$ ; communication,  $r = .42$ ,  $p = .001$ ), but the correlation for cognitive complexity was much lower and was not significant. In other words, the lead of high-school high achievers extends only through formal writing and the rhetorical aspect of communication, and does not help to predict complexity of ideation. 'Clean' writing tends to be set in place as early habit, whereas college develops more variability in thinking *per se*.

### Gender

Women who come to Evergreen were higher achievers in high school than Evergreen men ( $t$ -test; women's mean high-school GPA = 3.14; men's, 2.86;  $p = .008$ ), and they are higher achievers on the combined index of composition, communication, and cognitive complexity (women's mean = .24, men's, -.17;  $p = .03$ ). (These index scores may be called 'averaged standard scores.' The index is an average of three standard scores, and so the unit of measure might be dubbed an 'average of standard deviations.'<sup>3</sup>)

Is the advantage which women bring from high school a more technical advantage in formal composition, or a more general advantage in rhetorical effectiveness and cognitive complexity? While women are a little better in formal composition, they move ahead especially in effective communication. Their average standardized composition score is .23, and the men's is -.26 ( $t$ -test,  $p = .01$ ), indicating the lead in technical composition. But their advantage in rhetorical effectiveness or communication is even greater: the women's average standardized score of communicative effectiveness is .35, and the men's, -.41 ( $p < .0005$ ). The two groups differ slightly and in the same direction, but not, at a statistically significant level, in cognitive complexity. Men, on the average, try to express ideas just as complex as those which women express; but the women tend to be more successful in communicating, especially in linking concrete evidence to whatever general point they wish to make.

---

<sup>3</sup>The combined index (1) combines the standard scores for the composition, communication, and cognitive complexity ratings, and (2) averages all four years, as explained previously. Because standard scores are used, the mean is theoretically zero. The range originally expressed standard deviations; but since the three standard scores have been averaged, the unit of measure has become an arbitrary one, an artifact of the means of constructing the index. Fictively and with a little humor about the way something arbitrary can sound objective and scientific, the units have thus been called 'averaged standard deviations.' They range from -1.82 to 2.18.

One naturally wonders if all women, or only some women, are advantaged - or, to put the question the other way around, if all or some men are likely to receive lower ratings. In particular, one might ask if high-school background (just considered) or age (next to be considered) are underlying factors which contribute to what appears to be a gender difference. Controlling for age, one sees that among traditional-age students - those graduating by age 23 - the women are definitely ahead of men on the combined index. In this subgroup, 33.3% of the women were rated below average, and 66.7% above; whereas for men, 80% rated below and 20% above (chi-square test;  $p = .001$ ). Among older students, the difference runs in the same direction but is not quite as large, and the results are not significant. Controlling for high-school grade point average, more students who graduated from high school with a GPA below B were also rated below average on the combined index, but the difference was not statistically significant. Among students with a high-school GPA above B, the contrast looks dramatic, with 71.4% of women being rated above average on the combined index, in contrast to 62.5% of men below average; but the number of cases is small and the results do not fall within the range of statistical significance ( $p = .09$ ). In general, one can say that a cluster of women who are young high-school high achievers are disproportionately responsible for the women's overall lead; but men do not cluster so tightly and cannot be said as a group to be significantly at risk.

Combining the three scales year-by-year, one can say that the gender difference holds over time, but the difference decreases, indicating that Evergreen education, or general maturation, tends to ameliorate the initial freshman-year group difference in which women were significantly ahead. The difference in combined composition, communication, and composition scores is greatest in the freshman year ( $t$ -test; difference = .51 in favor of women;  $p = .004$ ) and is also statistically significant in the junior year (.35;  $p = .04$ ). The sophomore and senior differences move in the same direction, but are not statistically significant. Through the four years of college, men tend to catch up with women: the male-female difference declines from .51 to .21, in terms of averaged standard scores, from the freshman to the senior year.<sup>4</sup> This looks like a 'good,' in contrast to a gender-biased, outcome, if the goal of a liberal education is to overcome initial differences of race, class, and gender and to create a graduating group in which no demographic grouping is disadvantaged. From another standpoint, however, one could ask if men are getting more help, as they catch up with the higher standard set initially by women.

### ***Age at Graduation***

Perhaps the most surprising result regarding group differences is the strong performance of younger students and the *negative* relationship between age and combined composition, communication, and cognitive complexity ratings. Popular lore based on nationally declining SAT scores suggests that younger students might not do well as older. But if this is so, it cannot be substantiated in the present study, in which younger or traditional-age students - those arriving at college by age 19 and graduating

---

<sup>4</sup>Preliminary results from Fairhaven College (M. Eaton, personal communication) point in the same direction, except that in Fairhaven's summative evaluations, the males, who had consistently been rated below females in cognitive development, finally achieved a higher mean. But whether or not the difference is statistically significant has not yet been determined.

by 23 - more than held their own. Perhaps the allegations would hold up in other colleges, or in long-term longitudinal studies, or in a population of transfers as well as indigenous students. But the present data say that among Evergreen indigenous graduates in recent years, high-school directs and traditional-age students are doing very well, and if a group is at risk and in need of help with writing, it is their elders.<sup>5</sup>

The basic finding about age at graduation is a moderate negative correlation with the combined index ( $r = -.34$ ,  $p = .002$ ). This correlation accounts for 11.4% of variation on the combined index and is thus almost as large a contributing factor as high-school preparation. High-school preparation and age at graduation from college are partly overlapping but mostly independent factors. One might guess that older college graduates were people who did not do so well in high school, and thus deferred college; but this was only slightly, not markedly, the case. A tiny negative correlation hints at a link between high-school GPA and age at college graduation ( $r = -.15$ ), but it falls just short of statistical significance ( $p = .06$ ).<sup>6</sup>

One common preconception about older students does not apply here: these students are not primarily older women who deferred college for marriage and then matriculated later in life. This generalization might apply to older transfers, but not to the indigenous graduates. Of students over the median age of 23.5 (and thus over the traditional age of college graduation), only 38% are women, and 62% men. Moving up to age 30 and above, 30% are women, 70% men. At 35 and above, 44% are women, 56% men. Only at 40 and above (four women and three men) do women outnumber men among the older students. As mentioned before, women tend to receive higher ratings than men, and younger women were the 'point' group in women's higher ratings. But older students' ratings do not cluster consistently by gender. When sorted at every five-year interval from 25 to 40, both genders tend to fall further and further below the mean, with no significant difference between women and men.

A look at the correlation between age at graduation and combined ratings year-by-year generates a finding which bears upon need for writing instruction: the difference by age persists and even increases. The correlation between age and the combined ratings moves up from  $-.26$  in the freshman year ( $p = .02$ ) to  $-.37$  in the senior year ( $p = .001$ ). The size of this difference between older and younger students also increases: it is greater for upper-division students than for lower. It reaches  $.37$  in terms of averaged standard scores in the junior year ( $p = .03$ ) and is almost as large ( $.33$ ), but is not quite statistically significant, in the senior year ( $p = .09$ ). In other words, the age difference is *not* equalized over time by the overall effect of Evergreen education.

If a target group for special help with writing were to exist within the College, it would contain older students. But *most* older students are *not* at risk; despite the

---

<sup>5</sup>This tendency does not correspond with results at Alverno College (Mentkowski, Moeser, & Strait, 1983) regarding older students' cognitive development, but it does agree with preliminary results at Fairhaven College (Eaton, Personal communication).

<sup>6</sup>Most of the older students graduated from high school after, not before, the onslaught of grade inflation, so one cannot say their grades are a little lower just because schools used to give lower grades.

findings mentioned so far, one cannot actually predict that older students will be rated low in writing and thinking skills, because variance is high ( $s.d. = .89$ ) and many older students still fall in what one might call the safe zone not far from the middle of the index. Thus, one can say simultaneously that (1) the age difference does not 'naturally' equalize over time, but (2) older students do not constitute a clear target population for writing instruction. Some need help and some don't, but not the whole group across the board.

Who, then, are the lower-rated older students, and where do their difficulties lie? Addressing the first question, one can sort the younger and older students at ages progressively further from high school. With the line dividing younger from older set at age 23 (i.e., with matriculation by age 19, the traditional or high-school-direct age), the difference in average rating for the two groups is not quite significant; but moving up to ages 30, 35, and 40, the difference is significant and increases at each level. Students under 30 have averaged standard scores of .15; those over 30, of  $-.41$  ( $p = .05$ ). Moving up another level, students under 35 average .14; over 35,  $-.87$  ( $p = .02$ ). Moving up again, students under 40 average .13; over 40,  $-1.24$  ( $p = .006$ ). This means that a few students who are quite a bit older are dragging their group's ratings down considerably, as can be seen in the following plot, in which the horizontal line represents the center of the index and the vertical line designates age 35 (Figure 3, opposite).

Distribution seems almost random on the left side of the plot, up to age 35. But most notably, the upper-right quadrant is virtually empty, and the lower right, containing students over 35 whose combined ratings fell below the mean, is populated. Younger students received a mix of high and low ratings, but low scores belonged disproportionately to students who were much older than average.

Examining the components of the index yields a further point: older students differed from younger on all three components of the index, but most seriously on communication (less seriously with composition, and least seriously with cognitive complexity). The correlation with the communication scale is  $-.36$  ( $p = .001$ ). The correlation with the composition scale is significant but smaller ( $r = -.24$ ,  $p = .01$ ). The correlation with cognitive complexity is very small and not statistically significant. It is not, then, want of either formal rules or complex ideation which distinguishes older students; it is a matter of having something to say which the reader can follow, and of conveying it by connecting the general and the particular. In other words, the writing of older students tends slightly more than that of younger students to break down into generalities and concrete particulars.

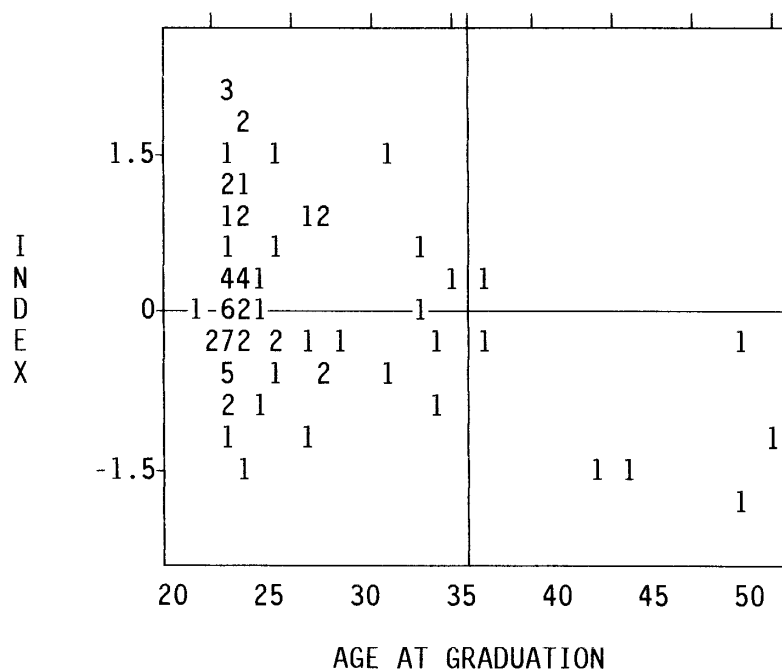
All of these data raise the question of *why* age relates slightly negatively to the combined writing and thinking index. (The question pertains, as always, to *indigenous* students, since the situation might be different for transfers.) Perhaps the greatest surprise is the strong performance of the youngest students. In other words, perhaps one can explain the situation in terms of their strength, without dwelling on the corollary of older students' relative difficulties. On the positive side, Evergreen is attracting high-school-direct freshmen who are able to hold their own against older peers.

Looking at the negative side, no one explanation of why older students might not do so well can stand alone, but several might work together. First, some of the

Figure 3:

AGE AND THE COMBINED INDEX:

PLOT OF POSITION ON COMBINED INDEX WITH AGE AT GRADUATION



Notes:

80 cases plotted.

Regression statistics of INDEX on AGE AT GRADUATION:

Correlation,  $-.33736$ ; R Squared,  $.11381$ ; S.E. of Est.,

$.82284$ ; Sig.  $.0022$

Intercept(S.E.)  $1.18346$  ( $.37019$ ) Slope(S.E.)  $-.04422$  ( $.01397$ ).

older students did not do well in high school; they number among the students with weaker academic backgrounds. Second, without college degrees, older students have been working in jobs which probably have not kept the fruits of earlier education fresh in mind, and probably have not been intellectually challenging. Third, since a few of the very oldest students had the most difficulty, one might speculate that habits of communication and cognition become, at some point in the life cycle, harder to change. And finally, perhaps the oldest students, when they do get to college, have trouble asking for or accepting help with their writing from younger writing tutors and younger or age-peer faculty.

Only the last of these factors - availability of tutors with whom older students need not feel embarrassed to work - is easily amenable to change. Overall, the point is that some (not all) older students need help with writing skills, especially at the level of rhetorical effectiveness, or of 'having something to say' and getting it across by linking the general and particular. They cannot be targeted as a closed group, since many of them have no writing problems, but help could be available to them in a form they can welcome. And this help should be forthcoming after the freshman year, because their ratings decline in relation to their younger colleagues especially in upper-division studies.

### ***Modes of Study***

One more relationship obtains between academic variables and ratings on writing and thinking, but it is quite small and hard to interpret. At least at first glance, it looks like students can take 'too much' coordinated studies, from the standpoint of improvement in writing and thinking, whereas students who enroll in more units of group contracts tend to do just a little better in terms of the skills or qualities measured in this study.

The basic findings are first, a small but significant negative correlation ( $r = -.25$ ,  $p = .03$ ) between units of coordinated studies and the combined index; and second, an equally small but significant positive correlation ( $r = .26$ ,  $p = .02$ ) between units of group contracts and the index. One might infer that a student can take interdisciplinary coordinated studies to a point of diminishing returns, so that large, broad-themed programs begin to inhibit concentration and rigor, whereas smaller, more disciplinary programs promote better writing and thinking. Any individual student might of course have a good reason to take one kind of program or the other in any given year on other grounds - especially the program's or contract's subject matter; but at a very general and aggregate level, one might become concerned about juniors and seniors taking a third and a fourth year of Core or entry-level interdisciplinary studies. An advanced student may find the writing too easy, when swimming in a school of freshmen.

To explore this possibility, the first step is to locate points of diminishing returns for coordinated studies, and of accruing benefit from group contracts. For coordinated studies, the difference in combined ratings is not significant between two and three years; but for students who went beyond three years of coordinated studies, the index ratings dip significantly. Of 80 cases, the 68 students who took up to three years of coordinated studies earned an average standard score of .17, while the 11 students who went beyond three years scored -.56. (This difference is significant at the level of  $p = .004$ .) For group contracts, the significant point is whether students worked in this mode for more than a quarter. Those who took less than 16 units of group contracts had average standard scores of -.42, and those who took more than a quarter, .30 ( $p < .0005$ ). Using these cut-off points, one finds significant differences in average standard scores which are related to mode of study.

What benefits to writing and thinking might accrue in group contracts? In terms of the components of the index, the correlation with group contracts depends on composition ( $r = .25$ ,  $p = .01$ ) and cognitive complexity ( $r = .23$ ,  $p = .02$ ), but not

communication. The benefits could arise from stronger motivation, or from writing more, or from working at a more advanced level; but probably not from receiving writing instruction *per se*, which is not often a feature of advanced group contracts. The sheer size of coordinated studies programs means they must embrace a greater range of student skills and backgrounds. Whereas more coordinated studies programs are lower division than upper division, group contracts tend to be the reverse. Some group contracts have prerequisites. The disciplinary spread of coordinated studies programs is usually broader, with group contracts usually confined to one or two disciplines. And so on the average, group contracts are more advanced in content than coordinated studies programs. As the Writing Across the Curriculum movement maintains, students probably learn the most about writing while addressing a challenging subject matter.

The flow of causality between index ratings and units of group contracts could move in either direction. One possibility is that group contracts, offering advanced studies and specialization, teach writing more effectively. But another is that students who already write well are more likely to move on with confidence to advanced work. Less advanced or less competent students are more likely to seek out easier programs.

A student's statement distills the motivation which can be aroused in a group contract to play for high stakes and take the work especially seriously:

I entered [a group contract] with certain apprehensions. Never, at TESC, had I worked with such a small, directed group. After five quarters of studying with the basic level coordinated studies mode, here I was, diving into an advanced level group contract. This was a small and intensive situation: one professor, fifteen extremely competent and motivated students. Yet I had some background ... and my teacher's respect for my ability. Accepting this challenge was, upon reflection, exactly the right thing to do. I now have a far greater faith in myself as a student, a thinker, and a researcher.

To sum up, the results of this section are negative from the standpoint of locating target groups who are at risk, and positive in relation to the overall tendency of Evergreen education to ameliorate group differences over time. Faculty might keep in mind that on a very general and aggregate level, the 'ideal type' of student who might have more trouble with composition, communication, and cognitive development is an older male who was less successful in high school and is perhaps shy or avoidant about undertaking advanced studies. But this formula cannot serve to target groups or individuals for attention because there is so much variability within groups that the average student in none of them is especially at risk. The answer to the question of who is at risk is no group at all, though the oldest students come closest. No policy targeting specific groups is called for. Writing instruction is an across-the-board problem, not one confined to identifiable minorities.

## CONCLUSION

The underlying design of this study has been to link quantitative findings, the part that goes 'by the numbers,' with qualitative interpretations, the inferences one can draw from what students say. This report contains some eloquent conclusions, but they have been spoken already, by the students themselves. These were the words of the students who ranked in the highest decile on the combined index, and of those who excelled from the standpoint of highest gain. The former tended to do well all through college, while the latter did not begin with a head start, but progressed farthest from their own benchmarks or baselines. Some equally impressive (if less sustained) discoveries came to light in the 'breakthroughs into ideation' which appear in average students' evaluations, when they break the mold of just going to school and find something of their own to say. These words painted a full picture of what Evergreen students can accomplish in terms of richness in thinking and writing.

With the most important conclusions already spoken, the final step is to review the recommendations which have been suggested in passing, and to state some qualifications about the context - the *dialogic* context - in which they should be considered.

### Summary of Recommendations

1. Using self-evaluations for evidence - on grounds that they illustrate writing and thinking skills transferable from college to the rest of life - one has to say that while student writing improves a little bit after the freshman year, the improvement is unimpressive. On the one hand, we can virtually guarantee that Evergreen students tend on the average to become better writers by the time of graduation. But on the other hand, the increment is small and students tend to reach an early plateau in their writing and to stay there. Since Evergreen does not focus on writing instruction after the freshman year, it would seem in the face of this evidence to be worth trying. (Nothing in the study can prove it would work; the study can only point to a problem worth addressing.) As stated above, a more general approach to "writing across the curriculum," with emphasis on help for sophomores and upper-division students, would seem to be called for, if improvement in composition and communication, as measured here, are reasonable objectives of liberal arts and sciences education. This should seem especially worth trying when one realizes that all of the evidence in this study comes from indigenous graduates, those who attended Evergreen for all four years. Transfer students presumably need even more help with their writing after the freshman year.

2. Writing instruction won't go far enough if it focuses on the formal rules of Standard English. It has to extend further, towards organizing what one writes around a thesis, or at least around some organizing principles or a point of view. A thesis is not just a prosthesis, a mechanical device; it has to be substantive, it has to be an *idea*. Teaching writing involves teaching *thinking* - getting things organized around 'something to say.'



3. Some common but typically unsuccessful strategies of self-evaluation are prevalent, and these strategies may illustrate general habits of writing and thinking which will not transfer especially well to the so-called 'real world.' To some extent they are age-appropriate or developmentally appropriate modes of 'procedural knowing'; but to the extent that faculty can help students to move past them, students would probably profit. These strategies center on what one might call a bureaucratic address to writing and thinking (perhaps indeed to life, despite all of the Evergreen emphasis on alternatives). Writing consists too much of listing things, of engaging not in thinking, but in exhaustive enumeration. Rather than think very hard about matters of substance, average students use 'default strategies' to enumerate regularly occurring activities, as organized by the calendar. These are effective strategies for getting the job of evaluation over with, and they are perhaps transferable to modest bureaucratic employment in the world after college. But they do not suggest that the student is developing the habits of thought or the evaluative and deliberative skills called for in the world's more responsible positions. In teaching writing and thinking, faculty might both (a) encourage these kinds of writing and thinking when they are developmentally new, and (b) urge movement beyond them as soon as possible when they have become familiar. The method - if there is one in this area of fragile insights and inspirations - is to teach students how to use their own emergent ideas as central or structuring principles for their writing. While no formulas can work in this area, nevertheless it was possible to chronicle eight different kinds of breakthroughs which are, so to speak, on the tip of average students' tongues - breakthroughs into ideation which are accessible in what Vygotsky called the "Zone of Proximal Development." A strong argument was made in the section on "Teaching Writing with Substance" that encouraging just this sort of breakthrough, to the point where it might be used habitually as the thesis or organizing principle of what the student writes, is the 'golden opportunity' in teaching writing and thinking. Every instance will be different; but the many examples from average-rated and high-rated self-evaluations provide confidence that much is possible and much can be encouraged besides business-as-usual.

4. The numbing detail of average self-evaluations and the lack of a clear focus on personal achievement make one ask how well such evaluations function as transcript material. Another model is used at Fairhaven College (Western Washington University), where seniors use earlier self-evaluations to write a retrospective summary or 'summative' self-evaluation. Preliminary findings at Fairhaven suggest that writing such a document is an important developmental experience in its own right, with scores on cognitive complexity increasing noticeably over the earlier self-evaluations. Anyone who reads large numbers of student self-evaluations is likely to think that such a procedure at least merits experimentation.

### **Deliberation and Dialogue**

It might be said critically that this study has simply demonstrated the obvious, or what everyone knew already. Didn't we already know that good writing has a thesis, or at least an underlying structure and organization? That formulating a thesis is not a mechanical matter, but a matter of finding something to say? That writing in particular, and education and every other 'calling,' tends to be successful to the extent that one involves the self?

But the same point can be offered constructively: if these are not striking theses from the standpoint of knowing, perhaps they are more important from the standpoint of acting, and will force some practical consideration of what is to be done, in particular about teaching writing after the freshman year. If this study demonstrated the obvious, then the obvious is now on display, for all to see and discuss pointedly. It is perhaps a time to act in relation to what we see.

But in orienting this study toward action, a strong qualification is in order: the kind of action called for is not just administrative action, but full discussion by the faculty. This report should not be used to justify any sort of change without dialogue. The reasons given here to justify the recommendations are not sufficient in themselves; rather, they are only the reasons which emerge *from the perspective of this study*. The context of writing and thinking, and of Evergreen education, is much larger than was researched here. Take the example of summative self-evaluation: this study led to such a recommendation because so many evaluations bore the reader with an overkill level of concrete detail. But this suggestion should not be instituted administratively because this study 'said so.' Rather, a whole range of issues merit consideration: was the finding simply a result of researcher fatigue - are allegedly boring average evaluations actually rather interesting documents, if one doesn't have to read a whole stack of them? Are self-evaluations as they stand perfectly adequate or even 'true' pictures, warts and all? Would summative evaluation weaken, or would it enhance, the important pedagogic procedure of writing evaluations at the end of every program? What procedures and resources would be required? What experience can be shared by those who have used this procedure and pedagogy at Fairhaven? These are matters for faculty discussion, to bring to bear a range of experience extending far beyond the present study.

If, then, this is a statement of the obvious, it is also a call for action; but its area is educational policy, which should be determined through deliberation by the faculty.

## REFERENCES

- Arendt, H. (1977). *Eichman in Jerusalem*. New York: Viking Penguin.
- Belenky, M., Clinchy, B., Goldberger, N., & Tarule, J. (1986). *Women's ways of knowing: The development of self, voice, and mind*. New York: Basic Books.
- Bruner, J. (1986). *Actual minds, possible worlds*. Cambridge, MA: Harvard University Press.
- Cederblom, J., & Paulsen, D. (1991). *Critical Reasoning* (3rd ed.). Belmont, CA: Wadsworth.
- Council of Presidents and State Board for Community College Education (1989). *The validity and usefulness of three national standardized tests for measuring the communication, computation and critical thinking skills of Washington State college sophomores: General report*. Bellingham, WA: Western Washington University Office of Publications.
- Educational Testing Service (1990). *Rhetorical effectiveness scoring guide for reflective essay*. Emeryville, CA: ETS.
- Finkel, D., & Arney, B. (1992). *The paradox of pedagogy*. Olympia, WA: The Evergreen State College (prepublication edition).
- Freire, P. (1970). *Pedagogy of the oppressed*. New York: Continuum.
- Gardner, H. (1991). *The unschooled mind*. New York: Basic Books.
- Gilligan, C. (1982). *In a different voice: Psychological theory and women's development*. Cambridge & London: Harvard University Press.
- Hunter, S. (1990, November 12). *An assessment of writing skill development based upon student self-evaluations*. Memo to Assessment Study Group, The Evergreen State College.
- Jung, C. (1921/1971). *Psychological types*. Vol. VI of *The collected works of C. G. Jung*. Bollingen Series XX; Princeton: Princeton University Press, 1971.
- Kegan, R. (1982). *The evolving self: Problem and process in human development*. Cambridge, MA: Harvard University Press.
- Knefelkamp, L. L. (1974). *Developmental instruction: Fostering intellectual and personal growth in college students*. Unpublished doctoral dissertation, University of Minnesota.

- Lakoff, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- Lakoff, G., & Johnson, M. (1980). *Metaphors we live by*. Chicago & London: University of Chicago Press.
- Lederer, T., & Hofvendahl, D. (1991, December 15). The Evergreen State College student evaluations: Some notes and thoughts. Memo to Steve Hunter, Office of Research and Planning, The Evergreen State College.
- Levinson, D. (1978). *The seasons of a man's life*. New York: Knopf.
- MacGregor, J. (1987). *Intellectual development of students in learning community programs 1986-87*. Olympia: Washington Center for Undergraduate Education, Occasional Paper No. 1.
- Michels, R. (1915, reprinted 1959). *Political parties*. New York: Dover Publications.
- Mentkowski, M., Moeser, M., & Strait, M. (1983). *Using the Perry scheme of intellectual and ethical development as a college outcomes measure: A process and criteria for judging student performance*. 2 vols. Milwaukee: Alverno College Productions.
- Moore, B. (1992, June). Student writing skills in Washington higher education. *Washington Assessment Group Newsletter*, 1 (6), 1-2.
- Moore, W. (1982-86). *The Measure of intellectual development: A brief review*. (Originally 1982; with periodic revisions, through 1986.) Farmville, VA: Center for Applications of Developmental Instruction.
- Moore, W. (1986). *Understanding students-as-learners through the Perry scheme: A rater training workshop*. Farmville, VA: Center for Applications of Developmental Instruction.
- Moore, W. (1989). *The measure of intellectual development: An instrument manual*. Farmville, VA: Center for the Study of Intellectual Development.
- Nelson, C. (1988). Varieties of critical thinking in science: From data to dialectics. Workshop materials presented at The Evergreen State College.
- Nelson, C. (1989). Addressing process and content: Teaching moves based on Perry and *Women's ways of knowing*. Workshop materials presented at conference on "Conflict over content: Knowledge, values, and higher order inquiry," University of Chicago.
- Perry, W. (1970). *Forms of intellectual and ethical development in the college years: A scheme*. New York: Holt, Rinehart & Winston.

- Perry, W. (1981). Cognitive and ethical growth: The making of meaning. In A. Chickering & associates, *The modern American college*, pp. 76-116. San Francisco: Jossey-Bass.
- Thompson, K. (1989). *Narrative assessment in collaborative learning environments: Encouraging connected and constructed knowing*. Paper presented at the AAHE Conference, Chicago, IL. Olympia: Washington Center for Undergraduate Education, Occasional Paper No. 2.
- Thompson, K. (1991). *Learning at Evergreen: An assessment of cognitive development*. Olympia, WA: Washington Center for Undergraduate Education, Monograph No. 1.
- Weber, M. (1958). *From Max Weber: Essays in sociology* (H. Gerth & C. W. Mills, Eds.). New York: Oxford University Press Galaxy Book.

**Appendix I:**

**"THE EVERGREEN STATE COLLEGE ESSAY READING"**

(Lederer & Hofvendahl, 1991)

**4 - Very good to excellent**

*Writer clearly demonstrates competence in writing on both the rhetorical and syntactic levels.*

**Content/Organization:** The writer

- clearly establishes a controlling idea
- provides sufficient and relevant supporting details which work in a coordinated way to support a central purpose
- progresses to a clear conclusion or ending
- demonstrates careful thought about the topic and often shows unusual insight or perspective

**Style/Voice:** The writer

- is sincere and candid, writing knowledgeably (possibly from personal experience) so that the overall effect is individualistic
- uses a tone appropriate to topic and audience
- consistently demonstrates syntactic variety

**Conventions:** The writer has a strong grasp of standard written English, and errors, if present, are minor.

**3 - Good**

*Writer demonstrates competence in writing on both the rhetorical and syntactic levels.*

**Content/Organization:** The writer

- establishes a controlling idea
- provides relevant supporting ideas
- provides an adequately stated conclusion
- demonstrates some insight or perspective

**Style/Voice:** The writer

- makes an honest attempt to deal with the topic with some sense of personal involvement
- uses a tone which is generally appropriate to topic and audience
- demonstrates some syntactic variety

**Conventions:** Writer generally demonstrates ability to handle conventions of standard written English.

## **2 - Adequate but undistinguished**

*Writer demonstrates adequacy in writing on both the rhetorical and syntactic levels.*

**Content/Organization:** The writer

- presents a controlling idea although it may not be readily evident
- provides minimal supporting details
- makes some attempt at a conclusion, which may be perfunctory
- demonstrates conventional and possibly mundane insights

**Style/Voice:** The writer

- attempts to deal with the topic but often relies on generalities, conventional truths, and predictable responses
- employs a tone which is bland and undistinguished
- demonstrates adequate but undistinguished facility with syntax

**Conventions:** The writer may make common errors in standard written English, but they are not serious or frequent enough to distract or confuse the reader.

## **1 - Poor (perhaps in need of remediation)**

*Writer demonstrates or suggests incompetence in writing, and the writing is flawed on the rhetorical and/or syntactic level.*

**Content/Organization:** The writer

- fails to present a controlling idea
- provides inadequate details
- fails to present a clear conclusion or leaves the reader in doubt as to writer's purpose
- demonstrates a lack of insight or the inability to go beyond the superficial in thinking and presentation

**Style/Voice:** The writer

- demonstrates little ability or willingness to deal with the topic on a personal or original level
- uses a tone which, if identifiable, may be inappropriate and/or indicates no sense of audience
- relies on primer prose and/or a series of short, choppy sentences with minimal modifications

**Conventions:** The writer has serious problems employing the conventions of standard written English.

**Appendix II:**

**"RHETORICAL EFFECTIVENESS SCORING GUIDE  
FOR REFLECTIVE ESSAY"**

(Educational Testing Service, 1990)

Reflective essays derive from the personal experience of the writer. But beyond the description and narration involved in communicating that experience, reflection requires probing into what this experience can show about the writer's life in particular, and, more importantly, about the writer's ideas of life in general.

The writer of a reflective essay works to see connections between experience and ideas, to test out thinking about an idea in the light of other experiences, and to arrive at new dimensions of the initial thinking.

Reflective essays are grounded in the concrete. An ordinary thing seen, read, or experienced triggers the reflection and leads to exploration of an idea. Occasions for reflections cover the range of personal experience from observations of natural phenomena, to recalling or witnessing events or to encountering a provocative idea in a novel or on the screen.

Unlike Observational Writing, which focuses on conveying one's personal perceptions, or Autobiographical Incident, which involves narrating an incident and evaluating its significance, reflective essays move to a different level of abstraction. The Reflective Essay ultimately identifies or discovers that its subject is, in fact, an abstraction - truth, beauty, patience, injustice. Reflective writers explore the meaning of this abstraction for themselves and for people in general. It is this people-in-general aspect of reflection - exploring the larger social implications of an idea - which is the hallmark of the Reflective Essay. It involves students in a unique kind of experience-based thinking.

The flow of thinking in exploring an idea may take shapes such as these:

The writer may first present the occasion (narrate a full incident, describe an observation) choosing details and images carefully as a way to ground the reflection that follows. The reflection then moves off on its own, perhaps with some reference to the initiating occasion.

The writer may launch an occasion but then move in and out of it along the way, reflecting on the idea it suggest.

The writer may construct a web of related, often parallel experience that serve as the stimulus for reflection.

The writer may focus first on a single occasion and then draw associations between it and other related experiences which build to an ultimate idea about people or the world at large.

The writer may begin with an idea from a quotation, proverb, or general experience and test concrete personal experience against it, reflecting about how



each experience relates to the idea. The reflection is refined more fully with each example until the idea has been tested from various angles.

Whatever thought pattern emerges, the writer's reflections explore the meaning of the occasion beyond the personal to the general.

Students who lack experience reading and writing the Reflective Essay usually respond in limited ways to CAP prompts: (1) they fail to ground their reflections in concrete observations or personal anecdotes or do so only superficially and then write a conventional "expository" essay about the idea in the topic; (2) they narrate a relevant personal experience but then neglect to explore the idea it suggests or do so only briefly, often in a moralizing way; or (3) they write a meditation rather than a reflection, turning an idea over and over but not grounding their ideas in personal experience.

## **6 - Exceptional Achievement**

### **Occasion for Reflection:**

The 6 writer memorably presents the occasion for reflection (a thing seen, read, or experienced), often with the fine detail of the naturalist or autobiographer. Though it does not dominate the essay at the expense of reflection, the occasion is nevertheless presented in extended, concrete detail. Whether it be an anecdote or nature observation or literary text the occasion grounds the entire essay in concrete experience.

Writers of 6 essays may use such strategies as these to ground their reflections:

- describing an animal, object, or phenomenon using concrete language rich in sensory detail
- recording specific behaviors, properties, or actions, often using narrative strategies such as pacing, dialogue, movement
- citing a quotation - poetry, prose, proverb
- constructing a web of related, often parallel experiences that serve as the stimulus for reflection.

### **Reflection:**

In a 6 essay the reflection or the idea suggested by the occasion is exceptionally thoughtful and convincing. The reader is impressed by particular insights. The writer is clearly thinking freshly, originally, honestly about the idea and has left commonplaces and clichés behind. The reflection tends to be extended, reflecting a serious, almost tenacious, probing and exploring of the idea.

The reflection may include generalizations about the writer's personal experience (signaled by "I" and "my") or more abstract commentary (signalled by "people", "they" or the editorial "we") about the idea and its broader implications. (For brevity's sake, these two types of reflection can be referred to as personal reflection and general reflection.)

Personal reflection may be understood as the first step away from narration of personal experience toward the Reflective Essay's characteristic idea abstraction of general reflection. Most essays scored 6 will have some explicit, insightful general reflection. In some notable papers, however, the writer's presentation of the occasion is at the same time a reflection. In these papers, the general reflection is implicit, embedded in phrases or clauses that cue the reader to move beyond the specific occasion to the abstraction that underlies it. The tone, established by a distancing of self from occasion, clearly conveys the reflective nature of such essays.

Though exploratory, the reflection seems to find a direction and reveal discovery or deepening insight, sometimes expressed as wonder, without a sense of conclusiveness.

### **5 - Commendable Achievement**

#### **Occasion for Reflection:**

Like a 6 essay, the 5 essay presents an extended concrete occasion. The occasion does not dominate the essay at the expense of reflection. A 5 essay lacks only the vividness and impact of a 6.

#### **Reflection:**

The writer engages in extended, thoughtful reflection. As in a 6 essay, the writer includes at least some general reflection. The occasional 5 essay will establish a reflective tone by an effective distancing of self from occasion. The 5 essay, however, will not carry it through so conclusively. The personal reflection and general reflection are serious and honest, but lack the intellectual leaps and freshness of a 6 essay. The essay reached beyond obvious statements about the occasion and idea. The reflection is not entirely predictable.

A 5 essay reveals direction or purpose but without the growing insight of a 6.

### **4 - Adequate Achievement**

#### **Occasion for Reflection:**

The 4 essay presents a concrete occasion, but may lack the detail or specificity of a 5 or 6. May be a full-fledged incident that dominates the reflection. Strikes the reader as a strong and interesting occasion.

#### **Reflection:**

Reflection will indicate a serious attempt to explore the idea suggested by the occasion but may be less well grounded in that specific occasion. The reflection may be intelligent but predictable or commonplace. The connection may seem tangential. The writer may rely on personal reflections about the occasion but

will still include at least a brief general reflection. A 4 may be characterized by thoughtfulness rather than discovery.

### **3 - Some Evidence of Achievement**

#### **Occasion for reflection:**

Writer presents an occasion but it may either be brief or dominate the essay. What appear at first reading to be occasions may actually be examples chosen to illustrate an initial generalization.

#### **Reflection:**

May have a meandering, rather than purposefully exploratory, quality. May rely on personal reflection to the exclusion of general reflection.

Reflection will seem obvious or even superficial, often taking the form of moralizing.

Some 3 essays will offer only extended reflection about the idea in the prompt, with little grounding in an occasion. The essay may seem generally competent and the reflection may be as interesting as in a 4 or 5 essay, but the writer has not yet learned that reflection must be carefully grounded in an occasion.

Other 3 essays will begin with what sounds like the end point of reflection, a conclusion or generalization which becomes the stimulus for recounting one or more illustrative examples.

### **2 - Little Evidence of Achievement**

#### **Occasion for reflection:**

The occasion may be brief or it may dominate the essay. There may be no occasion or the essay may be all occasion with little reflection.

#### **Reflection:**

If there is an occasion, the reflection may be additive or unfocused. It may be very brief and simplistic. Some 2 essays may be extended personal or generalized reflections on a topic with no grounding at all in an occasion.

### **1 - Minimal Evidence of Achievement**

#### **Occasion for reflection:**

If there is an occasion, it will be very brief and devoid of specificity or concreteness.

**Reflection:**

There is no reflection. There may be brief and superficial attempts at definition or statements of opinion rather than reflection.

**0 - Inappropriate Response**

**Off Topic:**

A paper that is off topic will not be even tangentially related to the prompt.

Appendix III A:

FOUR 'EPISTEMOLOGIES'

(Perry, 1970, 1981; Belenky et al., 1986.)

POSITION 2, *INFORMATION* ("dualism" or "received knowing;" Freire's "banking"): The student's main concern is "what to learn." Knowledge is information from Authorities (teachers and texts), and can be classified basically as right or wrong, true or false. The student's job is to make sure she got the teacher's or the text's information straight, and to be able to repeat it.

POSITION 3, *PROBLEM-SOLVING* ("early multiplicity," or "separate, procedural knowing"): In some areas, the right answers are not yet known. The student's main concern is "how to learn," and she begins to see herself as a problem-solver in different disciplinary domains. The methods for problem-solving seem to exist separately and discretely, and the student experiences no concern, as yet, with relationships among methods and domains.

POSITION 4, *RELATIVISM* ("late multiplicity," or "subjective" and "connected knowing"): The student's main concern is "how to think." She encounters diversity of informed opinion, wonders whether or not different points of view can be reconciled, and becomes concerned about forming her own opinion. Authority in the singular becomes authorities in the plural; and authorities are observed to differ because their methods and assumptions (and particularly the determinants of gender, class, and culture) afford competing points of view.

4a, *Oppositional stance* (subjective, argumentative, and critical; closer to the masculine stereotype): Everyone's entitled to his own opinion; you can't call anyone wrong or let them call you wrong; relativism is the only truth. Trust your own gut, and then use critical reasoning defensively to undo other people's assumptions. (From this stance, students often generate "independent-like" thought: they have observed what clear conclusions would look like, and they offer likenesses of such assertions without yet knowing how to muster coherent supporting arguments.)

4b, *Adherent stance* (connected, cooperative, and caring; closer to the feminine stereotype): I'm trying to express my own opinion the way I see it - the way that's right for me. I grant the same personhood to others and try to appreciate all the differences among opinions, mainly in terms of the different determinants that make people think the way they do.

POSITION 5, *MODELS AND THEORIES* ("contextual relativism" or "constructed knowing"): The student undertakes "thinking about thinking" (or "meta-thinking"). While she can no longer believe in absolute or certain knowledge, nevertheless she can inquire by constructing tentative theories and models, some of which are demonstrably more appropriate than others to the context at hand.

Appendix III B:

"SCHEME OF COGNITIVE AND ETHICAL DEVELOPMENT"

(Perry, 1981)

- POSITION 1: Authorities know, and if we work hard, read every word, and learn Right Answers, all will be well.
- TRANSITION: But what about those Others I hear about? And different opinions? and Uncertainties? Some of our own authorities disagree with each other or don't seem to know, and some give us problems instead of answers.
- POSITION 2: True Authorities must be right, the others are frauds. We remain Right. Others must be different and Wrong. Good Authorities give us problems so we can learn to find the Right answer by our own independent thought.
- TRANSITION: But even Good authorities admit they don't know all the answers yet!
- POSITION 3: Then some uncertainties and different opinions are real and legitimate *temporarily*, even for Authorities. They're working on them to get the Truth.
- TRANSITION: But there are *so many* things they don't know the answers to! And they won't for a long time.
- POSITION 4a: Where Authorities don't know the Right Answers, everyone has a right to his own opinion; no one is wrong!
- TRANSITION: But some of my friends ask me to support my opinions with facts and reasons.  
(and/or)
- TRANSITION: Then what right have They to grade us? About what?
- POSITION 4b: In certain courses Authorities are not asking for the Right answer; They want us to *think* about things in a certain way, *supporting* opinion with data. That's what they grade us on.
- TRANSITION: But this "way" seems to *work* in most courses, and even outside them.
- POSITION 5: Then *all* thinking must be like this, even for Them. Everything is relative but not equally valid. You have to understand how each context works. Theories are not Truth but metaphors to interpret data with. You have to think about your thinking.